DATA QUALITY – A KEY TO SUCCESSFULLY IMPLEMENTING ECSS

THESIS

Craig A. Lane, Captain, USAF

AFIT/GLM/ENS/09-07

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GLM/ENS/09-07

DATA QUALITY – A KEY TO SUCCESSFULLY IMPLEMENTING ECSS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Logistics Management

Craig A. Lane, MA

Captain, USAF

March 2009

AFIT/GLM/ENS/09-07

DATA QUALITY – A KEY TO SUCCESSFULLY
IMPLEMENTING ECSS

Craig A. Lane, MA
Captain, USAF

Approved:

______///Signed///______________________	18 March 2009
Dr. Jeffrey A. Ogden (Advisor)	Date
______///Signed///______________________	18 March 2009
Dr. Martha Cooper (Reader)	Date

AFIT/GLM/ENS/09-07

Abstract

In response to the Expeditionary Logistics for the 21st Century (eLog21) campaign initiatives published in 2003, the United States Air Force (USAF) pursued the acquisition of technology to help transform its logistics processes. With process mapping complete and a proposed roll-out schedule, forward progress towards full implementation of the Expeditionary Combat Support System (ECSS) continues. As a key enabler to achieving eLog21 initiatives, implementing ECSS will help transform current USAF logistics business processes. Integrating more than 450 legacy systems, and with a projected end-state in excess of 750,000 primary, secondary, and tertiary users, ECSS is the largest enterprise resource planning (ERP) system implementation in the world.

While the driving force behind an ERP system implementation is exploitation of the numerous benefits associated with transforming business processes, there are several key challenges to address which can mean the difference between success and failure. Data quality is one critical factor in the successful implementation of any ERP system. It is a key to optimizing system performance while maintaining an uninterrupted and acceptable level of support to the war fighter. This research evaluates data quality, focusing on the completeness and consistency of the data, in selected USAF legacy systems. Specifically, this study identifies invalid entries in the source data and also compares item record data between source (D043A) and downstream client (SBSS). This analysis lays the foundation for developing an action plan to allocate resources in an efficient and effective manner to support cleansing the legacy system data prior to migration into ECSS.

AFIT/GLM/ENS/09-07

To my family, friends, and classmates, thank you for your unwavering support and patience throughout this endeavor.

Acknowledgments

I would like to express my sincere appreciation to my advisor, Dr. Jeffrey Ogden, for his insight and guidance during this research. I am also indebted to my reader, Dr. Martha Cooper, who endured my writing and provided valuable input from the very beginning of this work. The support you both provided is certainly appreciated.

Additionally, I must acknowledge the efforts of everyone who pushed me through the nuts and bolts of constructing this research as well as gathering the data to support it. Throughout the course of this work, I've been extremely fortunate to have had carte blanche access to the extensive knowledge of supply experts Lt Col Robert Callahan and MSgt (ret) Reed Wilson. I'd also be remiss if I did not thank Lisa Fitzgerald for extending her time, insight, connections, and most importantly, her encouragement in the bottom of the ninth inning.

However, my largest debt of gratitude is owed to the staff at Qbase™, specifically Franz Weckesser, David Judson, and Bob Kinney, who all selflessly volunteered their time, their software, and their exceptional data prowess to helping me consume this data elephant one bite at a time. Thank you.

Craig A. Lane

Table of Contents

Page

List of Figures

Figure Page

List of Tables

DATA QUALITY – A KEY TO SUCCESSFULLY IMPLEMENTING ECSS

I. Introduction

Overview

The operating environment of the United States Air Force (USAF) has evolved considerably over the past decade. At home on sovereign soil as well as abroad, the culture and the organization are marked by change. Budget and resource constraints drive the need for efficiency, process improvement, and innovation. Transformation has become the broad underpinning of a vision communicated throughout the military chain of command. Support of this necessary shift in culture permeates the Department of Defense (DoD) from the very highest levels.

> "The opponents of change are many, and its champions are few, but the champions of change are the ones who make history."
>
> George W. Bush
> Former President

The impetus for transformation across the USAF logistics community began with the development of a campaign known as Expeditionary Logistics for the 21st Century (eLog21). The eLog21 initiative is an overarching effort to transform Air Force logistics business processes, and to provide the framework which will promulgate information technology development, and subsequent refinement, to facilitate that transformation. The backbone of the eLog21 initiative is a strategic map formally labeled Logistics Enterprise Architecture (LogEA). LogEA is the single authoritative source for operational architecture, systems architecture, and the transformation plan which defines the future state of Air Force logistics. It provides the specific description and documentation of the current state (as-is) and the future state (to-be), as well as the strategy to transition from the former to the latter (Fri, 2007). The eLog21 campaign and LogEA set the foundation for the USAF logistics community of the 21st century.

Through the implementation of an Enterprise Resource Planning (ERP) system, many commercial companies have transformed their business processes and improved performance. Recognizing opportunities for improvement in both efficiency and effectiveness, the USAF has increasingly sought the knowledge and experience of these civilian entities, to leverage not only their best ERP practices, but also to glean valuable insight from their lessons observed. These ERP systems streamline the flow and sharing of information, and connect the cradle-to-grave processes across organizational components. They enable future planning based on real-time data and support robust trend analysis. In short, ERPs serve to vent the traditional organizational silos and integrate all functions across an organization. This integrated environment encourages all functions within the organization to work together, from the procurement of raw materials to end-product sustainment, which ultimately leads to significantly improved performance across the entire supply chain.

Generally, most organizational change has a negative stigma associated with it and can be riddled with various challenges. Implementing ECSS is a monumental undertaking for the USAF. The challenges facing this endeavor are exceptional. There are several widely known pitfalls which can make a successful ERP implementation difficult and elusive. This research specifically addresses data quality and provides a solid baseline from which the USAF and the system integrator can mitigate data quality issues as the transfer from legacy systems to ECSS occurs.

Problem Statement

Though tremendous progress has been made in developing ECSS, the path to a successful implementation remains uncertain. Senior leadership communicated a vision based on transformation and provided the framework to facilitate the change. The capital resources in excess of $700 million were provided and an ERP system was selected for implementation (Pugh, 2007). The USAF formed Integrated Process Teams (IPTs) consisting of subject matter experts (SMEs) who worked with Computer Sciences Corporation (CSC), the system integrator, to blueprint/map processes and identify user-requirements. As the USAF makes forward progress toward incremental release, initial operating capability (IOC), and incremental legacy systems deconstruction, the need to verify and validate the data in existing legacy systems is ever-present.

While technology provides the vehicle for transformation, it is only as useful as the data which feeds it. The amount of data involved in this transition is enormous. It is paramount to mitigate the risk posed by inaccurate data through identification, and the subsequent repair and/or acceptance of that data through efficient and effective resource allocation. It would be challenging to define the cost-benefit regarding this issue due to its immense size. However, quality data is a force multiplier and a priceless key to ensuring a smooth transition to ECSS. While there is a substantial cost associated with correcting pre-implementation data quality issues, the cost associated with a stifled implementation due to inaccurate data is considerably higher. This research helps identify the potential risk by assessing the current state of item record data quality. It also identifies data shortfalls as well as areas of focus for resource allocation to support data cleansing and risk mitigation.

Research Questions

1. How complete are item records?
2. How consistent are item records?
3. Where should resources be allocated to address data cleansing/correction?
4. What are the potential implications of these results?

Investigative Questions

1. What are the valid data character entries for the analyzed data elements?
2. What constitutes a complete record for the purpose of analysis?
3. What constitutes a consistent record for the purpose of analysis?
4. What constitutes a quality record for migration into the ECSS database?

Summary

As a key enabler to the initiatives of the eLog21 campaign, ECSS provides the means by which the USAF can realize the objectives defined by senior leadership. This study focuses on the importance of addressing data quality issues prior to implementing ECSS. The intent is to help identify and mitigate the risks associated with inaccurate data, thereby shaping the data environment for a greater probability of implementation success. By taking a proactive approach to address data quality issues now, senior leaders will be much better prepared to tackle challenges as they arise during operational implementation.

This chapter provided an outline of the overarching motivation behind researching the stated problem. Both the research questions and investigative questions were posed to frame the research. A brief overview of the structure for the remainder of this study follows. The second chapter provides a review of the literature by exploring a brief background of transformation, commonly shared views on the benefits and pitfalls associated with ERP implementation, the importance of ECSS to transformation within the Air Force, and finally the significant role of data quality with respect to the successful implementation of an ERP system. The third chapter outlines the research methodology used to capture data on the subject as well as the investigative questions which focus this study to help answer the research questions. The fourth chapter addresses the results of the data analysis derived from the study. Lastly, chapter five states the assumptions and limitations of this research. Additionally, conclusions are discussed as well as recommendations to help create an implementation environment prone to success. Chapter five also outlines potential areas for future research.

II. Literature Review

Overview

This chapter provides a review of the supporting literature which sets a foundation for the subsequent research. Before discussing the specifics of data quality with regard to ECSS and its implementation, it is important to understand the basis of transformation within the DoD, more specifically, within the USAF. It is of equal importance to understand what an ERP system is as well as the potential capability it brings to an organization while noting, however, that there is substantial risk involved. Transformation is the catalyst which motivated an ERP system implementation within the USAF. The Air Force solution, ECSS, is a key enabler of this transformation. Implementing ECSS is a complex and monumental endeavor. This ERP system will be the largest single instance in the world and its implementation warrants an in-depth review.

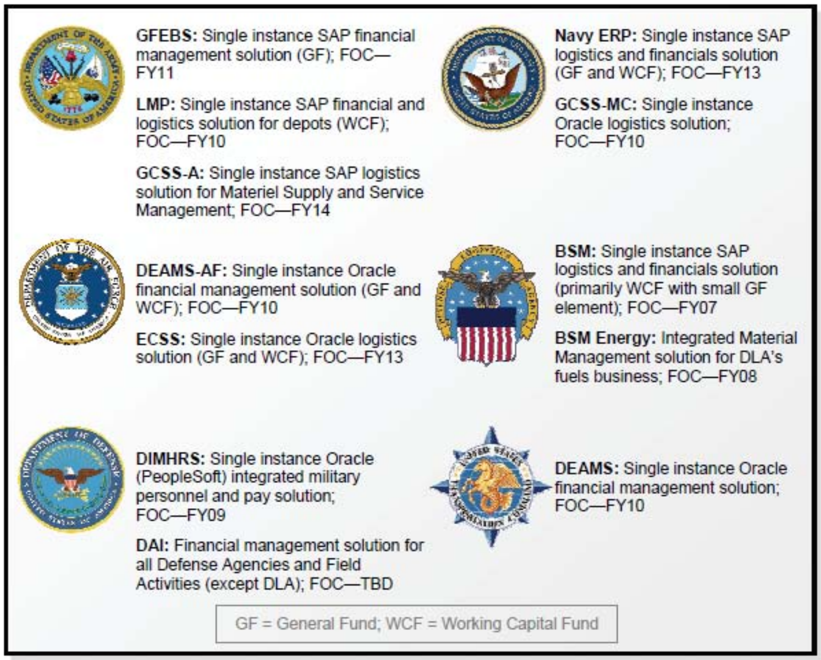
There are a multitude of potential benefits associated with the success of implementing ECSS. At the same time, there are a multitude of pitfalls which could impede that path to success. Data quality, in a broad sense of the term, is the core of this research, and will be defined for the purposes of supporting this research. Current Air Force guidance provides a basic knowledge of legacy system data requirements. This defines a framework for comparing the quantitative data collected from the selected legacy systems. The literature review will discuss commonly observed pitfalls as well as lessons observed through ERP system adoptions in both the DoD and the commercial sector. The lessons addressed are focused towards data quality and the significant role it plays, as either a vital bridge or a critical gap in the implementation process. Finally, this review will conclude with a brief discussion of the benefits of data cleansing prior to implementation.

Transformation

> "Just as we must transform America's military capability to meet changing threats, we must transform the way the Department works and what it works on. Our challenge is to transform not just the way we deter and defend, but also the way we conduct our daily business".
>
> Former Secretary of Defense Donald Rumsfeld

The United States military represents one of the largest and most complex organizations in the world. Since the end of the Cold War, the military mission has evolved and become increasingly dynamic. Long gone are the days of ample resources: robust manning, adequate capital, equipment, and infrastructure. The new face of war which has developed over the past decade has tested military limits on varying fronts, primarily in Iraq and Afghanistan. Additionally, involvement in humanitarian operations and military operations other than war, has also levied a significant impact on already scant resources. Weapons platforms suffer fatigue and extensive sustainment costs due to excessive use, while personnel right-sizing occurs as a trade-off to fund recapitalization efforts for these worn platforms. Budget constraints in light of the Global War on Terrorism (GWOT) have become the norm rather than the exception. Some of the driving factors over the past decade leading up to this point include a 50 percent increase in personnel cost despite manpower reductions, and an increase in aircraft fleet operations and maintenance costs by 87 percent. Additionally, DoD and Air Force budgets continue to steadily decline (Tew, 2006). As a result, senior leadership recognized the need to drive efficiency into military processes in an attempt to prosper in a resource-constrained operating environment. These leaders looked at successful commercial organizations and realized that the military could potentially benefit by adopting commercial industry best practices.

While each of the individual services adopted their own transformation initiatives, Air Force leadership introduced Expeditionary Logistics for the 21st Century (eLog21). The eLog21 initiative leverages the latest technologies to enable the Department of Defense (DoD) and Air Force logistics visions, while driving cost down through efficiencies gained by implementing industry and Air Force best practices (DAF, 2003). When fully realized, eLog21 will have transformed and enhanced business processes across the entire AF logistics community. Embedded within the eLog21 initiatives is a strategic road map, or Logistics Enterprise Architecture (LogEA), which shapes the transformation. The structure of LogEA revolves around the Supply Chain Operations Reference (SCOR) model. It outlines the current state of the Air Force logistics community, as well as the intended future end-state. This architecture provided the framework for selecting and subsequently implementing an Enterprise Resource Planning (ERP) system to meet Air Force logistics requirements, and to transform business processes horizontally as well as vertically. ERP implementation is about business transformation, not technology (Coker, 2006). Figure 1 outlines target ERP programs designed to transform business processes across the DoD; however ECSS will provide the vehicle to drive the transformation of the logistics enterprise across the USAF.



(U.S. DoD, 2007)

Figure 1 – DoD's Target ERP Programs

Enterprise Resource Planning (ERP)

A Google® search for "definition of ERP" returned approximately 245,000 results. It is obvious these definitions share several common key words and phrases such as "integration", "multi-module", and "amalgamation of processes". One of the more thorough definitions discovered is from BusinessDictionary.com, which states an ERP is an:

> "accounting oriented, relational database based, multi-module but integrated, software system for identifying and planning the resource needs of an enterprise. ERP provides one user-interface for the entire organization to manage product planning, materials and parts purchasing, inventory control, distribution and logistics, production scheduling, capacity utilization, order tracking, as well as planning for finance and human resources. It is an extension of the manufacturing resource planning (MRP-II). ERP is also called enterprise requirement planning."

BusinessDictionary

In today's global market, many successful organizations have revolutionized their business processes and improved performance through the implementation of Enterprise Resource Planning (ERP) systems. A considerable amount of research case studies exist regarding both the successes and failures of the organizations which implemented ERPs. Because this paper is focused on a specific area of implementation, a review of all specific business process improvements is not in order. A brief review of Neway and DLA reveals some of the potential successes which can be achieved by adopting and implementing ERP technology. These examples provide perspectives from both the commercial and military sectors.

A study of Chinese valve manufacturer, Neway, published in 2008, is a prime example of how an ERP system can benefit an organization. Following implementation, Neway was able to recover approximately $20,000 annually in lost sales. A 15-day inventory reduction resulted in $1 million in annual savings and reducing the monthly purchase frequency from 50 orders to 10 orders saved $4,800 annually. Below, Table 1 summarizes the additional benefits experienced at Neway, only 6 weeks post-ERP implementation (Bose et al., 2008).

Table 1 – Benefits of an ERP Implementation at Neway
(Outbound Order Fulfillment and Inventory Metrics)

Operational measures	Pre-implementation	Post-implementation
Commitment to fulfillment	80%	98%
Average lead time	45 minutes	30 minutes
On-time delivery percentage	80%	95%
Average safety stock period	40 days	25 days
Inventory accuracy	85%	99%
Average monthly purchase frequency	50	10

(Adapted from Bose et al., 2008)

In 2002, the Defense Logistics Agency (DLA) began implementation of their Business Systems Modernization (BSM) ERP. As a core combat logistics supply agency for the DoD, they manage approximately 5.2 million supply items totaling roughly $18 billion in annual business. During the implementation of BSM, DLA's annual sales and services increased from $17 billion in FY 2001 to almost $35 billion in FY 2005. Despite nearly doubling their operations tempo due to the GWOT, DLA managed to continue their business transformation and realized significant results with the implementation of BSM (U.S. DoD, 2007). While it may take several years following fully operational capability to accurately capture all of the benefits of BSM, the short-term results are impressive. Table 2 provides a summary of the successes at DLA.

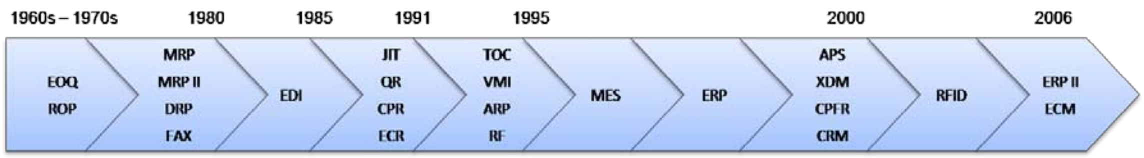
Table 2 – Benefits of BSM at DLA

	FY 2000	FY 2007
Cost of Operations	22.1%	13.1%
Average Order Processing Time	> 1 work day	< 4 hours
Overall Material Availability	88%	92%
End-of-Year Financial Close-out Time	2 weeks	1 day

(Adapted from U.S. DoD, 2007)

These examples are not intended to be representative of all ERP implementations. For every ERP success story with an organization, there is likely a tremendous ERP failure associated with another. These examples are presented simply to depict some of the potential benefits associated with the successful implementation of an ERP.

More than two decades have passed since the first documented ERP was implemented. The literature regarding the history of ERPs is also extensive. Only a brief synopsis will be addressed here as the focus of the research is not reliant on an in-depth knowledge of the entire historical timeline. Figure 2 depicts the evolution of ERPs. A detailed description of the listed acronyms can be found in Appendix A.

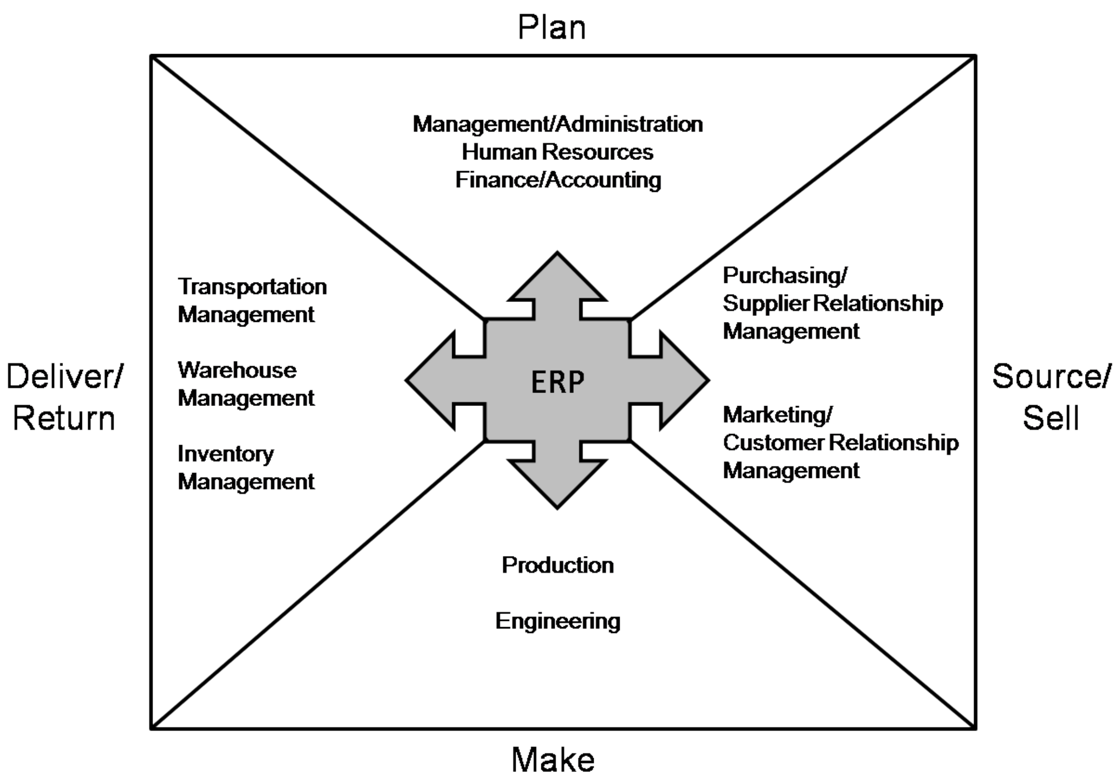


(Adapted from Fawcett et al., 2007)

Figure 2 – Evolution of ERPs

ERP systems evolved from Materials Requirements Planning (MRP) and MRP II systems dating back to the 1960's. These early systems were very narrowly focused and functionally aligned with regard to organizational stove pipes. They offered little, if any, intra- and/or inter- firm communication which led to inefficient and cost-inhibitive operations. As terms like "*transformation*" and "*supply chain management*" have been developed and applied to business processes over the past couple decades, IT has evolved to support and compliment these processes. Despite the Gartner Group coining the term "ERP" in 1990, Siemens company, in cooperation with SAP (a German-based software company), was the first to implement an ERP system in 1987 (Yu, 2005). ERPs today serve to streamline and standardize processes across the entire supply chain, ventilating the proverbial organizational silos and facilitating communication on a global scale.

Many business processes in the Air Force are disjointed. Information sharing is mediocre at best and duplicative processes are prevalent. Several hundred legacy systems contribute to a lack of both efficiency and effectiveness across the AF logistics enterprise. ERP systems are designed to tightly integrate the functional areas of the organization, and to enable the seamless flow of information within and across those functional areas. Effectively implemented ERPs centralize business process information and integrate processes to maximize performance (Lawrence et al., 2005). Figure 3 is a generalization of how ERPs centralize the business processes of the SCOR model. Details regarding the business processes can be found in Appendix B.



(Adapted from Fawcett et al., 2007)

Figure 3 – ERP-Centric Business Processes

ERP systems facilitate the flow of information to connect the cradle-to-grave processes across inter- and intra-organizational components. They enable future planning based on real-time data and support robust trend analysis, providing a more reliable source which leads to more informed and more accurate decision making.

Most implementation projects are unique in many ways, however, they all share several common issues, regardless of the system implemented. The overriding objective of most companies is to complete the project on-time and within the budgeted resources. It is safe to assume that the USAF would follow this line of thinking: on-time and within budget. In order to meet these objectives, ERP projects must be carefully planned and efficiently managed (Mabert and Venkataramanan, 2003). The USAF established both the ECSS Program Management Office (PMO) and the Logistics Transformation Office (LTO) to facilitate these objectives. The PMO is intended to ensure USAF requirements are met on-time and within the budget. The LTO gathers and consolidates USAF requirements, and acts as an advocate on behalf of the logistics community. The systems integrator, Computer Sciences Corporation (CSC), coordinates with both offices to execute the implementation of ECSS.

There is no industry standard defining success or failure with respect to an ERP system implementation. It seems generally accepted that success is defined by a combination of meeting projected budgets and implementation timelines, as well as process efficiencies and cost savings realized across the organization. Effectiveness of the ERP systems, post-implementation, is also a crucial indicator of success. The success of an ERP system is measured by its impact on technological, business, and human resource requirements. As with success, there is rarely a single identifiable flaw unique to failure; however, data quality can be a factor (Harris, 2003).

Despite the potential benefits associated with ERP systems, there is also considerable risk. Studies conducted over the past 10 years led to relatively dismal results. According to Trunik (1999), 40 percent of ERP systems perform to only some of their full effectiveness and 20 percent are scrapped as complete failures. While success is an attainable goal, it will not be easily achieved. Despite several existing inconsistencies with respect to system inefficiency and failure, it is both a logical and safe assumption that the implementation of ECSS will encounter several challenges leading up to implementation, as well as throughout its evolution and into its sustainment phase following implementation.

Expeditionary Combat Support System (ECSS)

ECSS is based on a commercial off-the-shelf (COTS) platform which will provide a single solution to integrate data from several hundred legacy logistics systems and drive efficiency into the logistics community (DAF, 2003). It is an Oracle-based platform supplemented by the Industrial and Finance System (IFS), which focuses on maintenance, repair, and overhaul; and ClickCommerce®, which focuses on advanced planning and scheduling. Together, these three information technology (IT) platforms comprise the Oracle Product Suite (OPS). This technology will facilitate data sharing across the entire AF logistics community from the procurement of raw materials to the finished product. The primary overall benefit is substantially improved support to the war fighting mission. Additionally, ECSS is expected to reduce inventories, reduce maintenance cycles, reduce administrative burdens, improve resource allocation with respect to demand, improve fiscal posture, and improve product and data quality. Specifically, realizing a 20% increase in equipment availability and a 10%, or $2.75 billion decrease in Operations & Sustainment costs by the end of FY 2011, are success bars set by USAF leadership (DAF, 2003).

Implementation of ECSS will occur in multiple phases. Five Integrated Process Teams (IPTs) were formed in alignment with the SCOR model (plan, source, make/repair, deliver/return, enable) to map or blueprint in excess of 1,000 logistics processes. When blueprinting is complete, the IPTs will work with the system integrator to perform a gap analysis, comparing the requirements of the logistics community with what the software provides, and then determining where software modifications are needed. The blueprinting phase will be followed by incremental legacy system deconstruction, fielding/release, data lifecycle management, and organizational change management. All of the phases overlap to some degree while others will be ongoing throughout the entire implementation process. The first operational test and evaluation is forecasted to take place in April 2010 (Pugh, 2008).

The enormity of implementing ECSS, to drive the transformation of AF logistics, is a huge undertaking. Inevitably, there will be barriers to success throughout, and after, implementation. It is imperative that senior leadership, current and future, get educated on the capabilities which ECSS can bring to the fight and to continually focus their subordinates on the ultimate, long-term benefits associated with this level of change. Change comes with a certain level of discomfort, however. It is incumbent on leadership to mitigate the effects of that discomfort and to promote a positive culture of acceptance and adaptation. The successful transformation of the logistics community depends heavily on the successful implementation of ECSS. The successful implementation of ECSS depends heavily on the organization adopting and embracing a positive attitude towards implementation and transformation. The mutually-dependent benefits of both the business process change, and the enabling technology, require unwavering support from across the organization. The reality, however, is that it will be quite some time, regardless of the measurement used, before anyone can determine whether the USAF achieved success or suffered failure.

Data Quality

"A great plan based on wrong information is doomed to failure" (Schumacher, 2007). Before discussing the importance of data quality with respect to ERP implementations, specifically ECSS, it is important to develop a foundation regarding the definition of data quality for the purposes of this study. It is also significant to note that there seems to be no strict industry standard for terminology. While they are not necessarily used interchangeably, the terms data cleansing, data integrity, data quality, data accuracy, and data management are used in similar contexts across varying literature with regard to data as a critical success factor (CSF) in ERP system implementation. For the purposes of this research, the term data quality will be used consistently in the context of the definition in the following paragraph.

In the absence of any industry standard, this research adopted a standard data terminology framework proposed by Dave Becker who is leading a developmental project called Air Force Inventory Data Quality Management (AFIDQM). The AFIDQM project focuses on several areas including data quality and its potential payoffs, enterprise data quality management strategy, and information manufacturing systems' inventory data. AFIDQM dovetails with this study as the research provides, to some extent, a proof of concept. While not the focus of this research, it does to some degree highlight the need for and benefit of standardized terminology with regard to data across the enterprise. His proposed framework follows:

- Quality – something fit for purpose
- Quality Data – data fit for its use
- Quality Data Characteristics –
    - Accurate
    - Precise
    - Complete
    - Consistent
    - Timely
    - Authoritative

(Becker, 2009)

Utilizing this framework as a reference for analysis, the scope of this research will focus on the characteristics of "complete" and "consistent". Completeness is defined as the degree to which data elements are present when/where they are required. Consistency is defined as the degree of freedom from variation or contradiction (Becker, 2009). The definitions for all six quality data characteristics in their entirety can be found in Appendix C.

In reviewing literature on the subject of data, it becomes potently evident that the significance of data quality to a successful ERP system implementation cannot be understated. It is critical not only to a successful initial implementation, but also to sustaining and exploiting long term operational effectiveness and efficiency. In a 2007 article, Emily Grantner stated, "A system is only as good as the data within that system. An increasing amount of organizations are discovering this as they upgrade older legacy systems into ERP systems" (Grantner, 2007:4). Several studies identify multiple CSFs that shape the successful implementation of an ERP system. Data quality is one key factor to ensuring success by providing a system operating with clean data. Quality data ensures smooth operations because end-users are more likely to trust and embrace a reliable system. Data quality is one of the most significant challenges facing successful implementation of an ERP system (Lawrence et al., 2005).

Since the development of early MRP systems which evolved into today's ERP systems, experts in the field recognized the importance of data quality. Effective and efficient system operations depend on the integrity of relevant data (Tersine, 1994). Tersine (1994) also noted that a lack of record integrity is a major reason for the failure of systems to live up to expectations. Furthermore, he states that computer-based systems, more so than manual systems, will not perform satisfactorily with poor files and records. In short, the output from a computer-based MRP system cannot be better than its input (Tersine, 1994).

More than a decade later, data quality is still held in critical regard. Sun et al., (2005) identified five CSFs with regard to an ERP system implementation. Data was prioritized number two in importance behind people, which included education, training, skills development, and knowledge management. Ngai et al., (2008) reviewed several ERP implementations across ten different countries and identified 18 CSFs. Although the CSFs were not rank-ordered, data management was included in the list of 18. With specific regard to inventory data accuracy, Titmuss (2001) estimated 80 percent of supply chain management problems could be traced to inventory records that are inaccurate. He also identified poor database accuracy as 1 of 12 reasons which consistently leads to ERP implementation shortfalls and/or failures (Titmuss, 2007). According to Caruso (2007), missing or inaccurate data can be a true project killer. This statement implies two distinct, but highly related issues; the absence of data as well as inaccuracies in existing data. Defined separately, or when combined, both lead to shortfalls in the system. Small data quality issues can quickly compound and grow into large issues. The effects can grow substantially across the system, especially in a system as large as ECSS. The new system must have clean data to start with or it will be handicapped from the moment it goes live (Lawrence et al., 2005).

Without quality data, i.e. data that both exists and exists accurately, the ERP will not function effectively and will not produce the results touted before implementation. The end-user will not trust the new system if they question the information it generates as a result of inaccurate data. Without user buy-in, the success of the implementation can be significantly hindered. A system already plagued with data quality issues will likely be doomed to failure because users abandon it. They will revert to using old, inefficient systems, or locally developed databases which they are comfortable and familiar with. The effects of this behavior across the enterprise can be severe. As the system loses credibility among users due to inaccurate output, it subsequently becomes unreliable for the organization.

Companies, who completed their ERP system implementation on schedule, as well as on, or under budget, shared several common characteristics including key technology issues. Data quality and technology infrastructure were addressed early. (Mabert and Venkataramanan, 2003). The DoD's Enterprise Transition Plan 2007 identified data cleansing as a key lessons learned from DLA's implementation of BSM.

> "Cleanse data up-front to ensure up-to-date, accurate, and authoritative information. This also reduces the amount of time spent designing interfaces to handle bad data."
>
> U.S. DoD ETP (2007: 22)

There is no question among academics relating the importance of data quality to the success of any ERP system implementation. It is consistently ranked among the top CSFs identified in most studies. There is also universal agreement on the fact that data quality should be considered early in any project of this type and that data cleansing is a must before going live with any new system.

While the focus on data quality tends to point towards successful ERP system implementation, there are other benefits associated with an operational system fueled by clean and accurate data. Reducing, and attempting to eliminate inventory inaccuracy can reduce supply chain costs as well as out-of-stock levels (Fleisch and Tellkamp, 2005). Over the past two decades, organizations have realized cost savings through reduced inventories and just-in-time supply functions. The quality of an item's stock record affects the accuracy of physical inventories. Inaccurate inventories drive up costs across the entire supply chain, regardless of whether the inaccuracy results in excess inventory sitting in a warehouse or it leads to a stock out situation at the base level. The real benefits and potential of these ideas can only be achieved with quality system data. This holds especially true for the USAF with regard to the successful implementation of ECSS. While allocating resources at the beginning of implementation to address data quality issues may be costly and time consuming, the long term benefits will be much more significant to the overall performance of the organization (Grantner, 2007).

Another important aspect concerning data quality is data management, which relates to system credibility among users and system reliability across the organization. This is a distinct and critical piece of the data environment, and it encompasses the attributes which directly affect data quality. Without effective data management, the quality of data within any system suffers. Data governance and operating policies should be established and enforced to ensure, or at least maximize, accurate data entry. This structure should also include sufficient means to identify and correct inaccurate data as well as serve to prevent recurrence. The subject of data management, including data quality, has remained a focus area over the past few decades with regard to ERP system implementations. According to Tersine (1994), “file integrity is not a one-time affair, but a constant vigil”. The relationships across the data environment are depicted below in Figure 4.

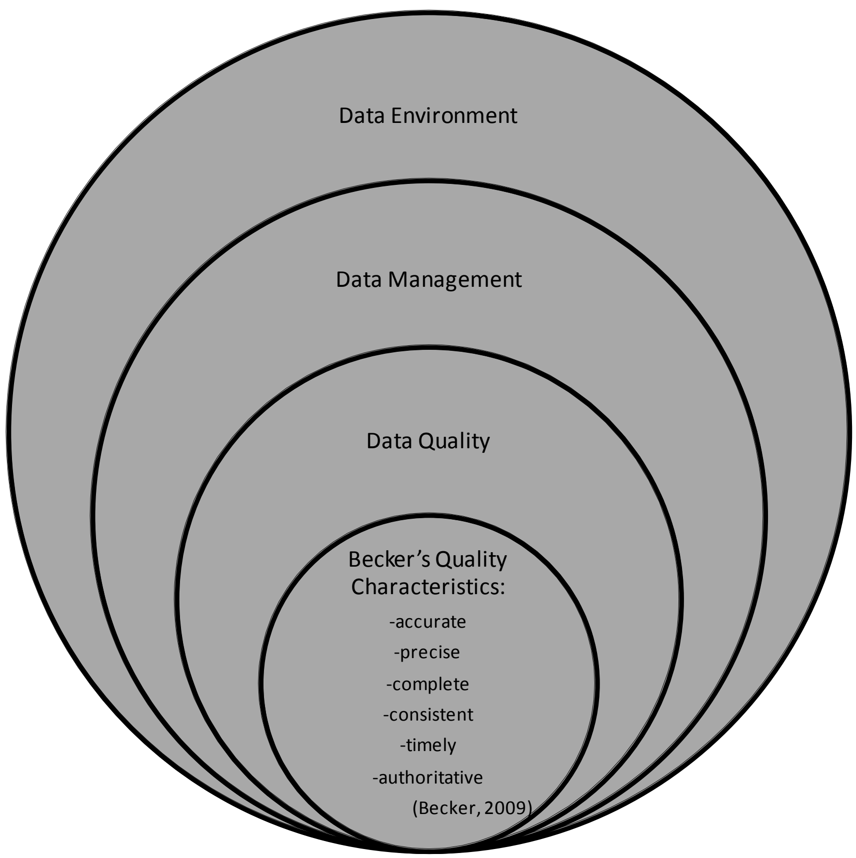


Figure 4 – Data Environment

Since ERP systems contain various modules which are intricately linked with each other, data should be managed properly to ensure accuracy (Ngai et al., 2008). Data management represents yet one more element of ERP system implementation which is of key importance. It shares ties with other critical areas such as organizational governance and policy. However, these subjects are all outside the scope of this research and as such, will not be further discussed.

Summary

Despite the importance given to the subject of data quality, the literature reviewed seems to be relatively devoid of any specific information defining what actions *should* be taken, or what actions *were* taken, to address this critical issue. This is also true when speaking in terms of data management or any of the other contextual terms identified earlier. While a multitude of case studies exist addressing ERP system implementations, none reviewed for this research outlined any specific actions taken in the realms of data quality and/or data management. As previously stated, all reviewed studies indicate academics universally agree that data, in a broad sense of the term, is a top CSF in any ERP system implementation. Several cases observed organizations using some form of electronic data interchange (EDI) to cleanse data before or during migration into their new ERP system. In most cases data quality was simply identified as a CSF for implementation. However, there was no indication of, or reference to, the specific actions taken when addressing the issue of data quality, hence the foundation of this research.

From a supply chain management viewpoint, there is general acceptance that the USAF is, in some ways, similar to civilian organizations, but in many other ways, quite different. As the USAF embarks on its journey to implement ECSS, it is not traversing uncharted territory; rather territory charted on a much smaller scale by sister services and several civilian entities. While forward progress continues, there is still a significant number of challenges ahead. The lessons observed through other organization's pioneering ERP implementations serve to lay a solid foundation from which the USAF can build upon. Despite the amount of importance it has earned, beginning with early MRP implementations, "data" remains a broad and ambiguous term defined by several related and smaller parts. Furthermore, data is a critical component touching other well defined areas across the entirety of any supply chain.

Past research doesn't seem to provide or define any detailed actions taken to attack the issue beyond using some form of EDI. Future endeavors, such as the International Organization of Standardization (ISO) 8000 series standards, seek to define data standardization for users around the world (Grantner, 2007). On a smaller front, the AFIDQM project is focused on setting a service-specific standard for several elements of the data environment (Becker, 2009). EDI alone will not mitigate this issue for the purposes of the USAF implementation of ECSS, nor is a future timeline going to be of use in trying to address a problem requiring near-immediate action. This study will address the data quality of item records by comparing base-level data from the Standard Base Supply System (SBSS) to the source data contained in the Master Item Identification Database (MIIDB). The intent is that the results of this study provide proof the utilized model is useful as a general guideline to focus data cleansing efforts in a resource-constrained operating environment, not only for the USAF, but also for those considering future ERP system implementations. Additionally, this research will help set the stage for future research serving to fill the literary gaps regarding what actions were taken as well as what actions should be taken to mitigate data quality issues prior to implementing an ERP system.

III. Methodology

Overview

This chapter addresses the selected research methodology, unit of analysis, research design, data sources and collection, and data analysis techniques employed. An experimental methodology, focused on quantitative statistical analysis was used to underpin this research. The research model developed to support this study is experimental in nature, as the review of the literature did not find a model to use as a guide. This model is intended to identify particular data elements within a data record which can be used to observe the relative quality of a population of like items. The quantitative aspect provides a current statistical snapshot of selected legacy system data. It provides a baseline from which inferences can be made with regard to data quality prior to legacy system deconstruction and data migration into ECSS. Before progressing with the remainder of this chapter, the research questions, proposition, and investigative questions are re-stated to provide context for the intent of this work.

Research Questions

These questions are designed to help keep the research focused. Data quality is a very broad topic and it overlaps several other key issues when discussing an ERP system implementation such as ECSS.

1. How complete are item records?
2. How consistent are item records?
3. Where should resources be allocated to address data cleansing/correction?
4. What are the potential implications of these results?

Dealing with several hundred legacy system databases is a monumental task. These research questions are broad so that they can potentially be applied to any system simply by changing the unit of analysis. They set the foundation for the investigative questions which follow.

Investigative Questions

1. What are the valid data character entries for the analyzed data elements?
2. What constitutes a complete record for the purpose of analysis?
3. What constitutes a consistent record for the purpose of analysis?
4. What constitutes a quality record for migration into the ECSS database?

The answers to these investigative questions aid in selecting the appropriate systems to sample. Additionally, they help to identify an appropriate unit of analysis as well as the specific data required for analysis. Coupled with the research questions, the answers to these investigative questions narrow the scope of this work, keeping the research focused and manageable.

Unit of Analysis

Air Force Manual (AFMAN) 23-110, Vol 2, Part 4, Ch 5, Table 5.1 lists all 225 types of data records found in the Standard Base Supply System (SBSS). For this research, the item record was selected as the unit of analysis. This is one of 225 different types of records residing in SBSS and contains 106 data elements. As defined in Vol 2, Part 4, Ch 7, Attachment 7A-2, Para 7A2.1, the item record contains sufficient data elements to manage most items. Separate records are maintained for all equipment and supply items on which accountability must be maintained (AFMAN 23-110, 2009).

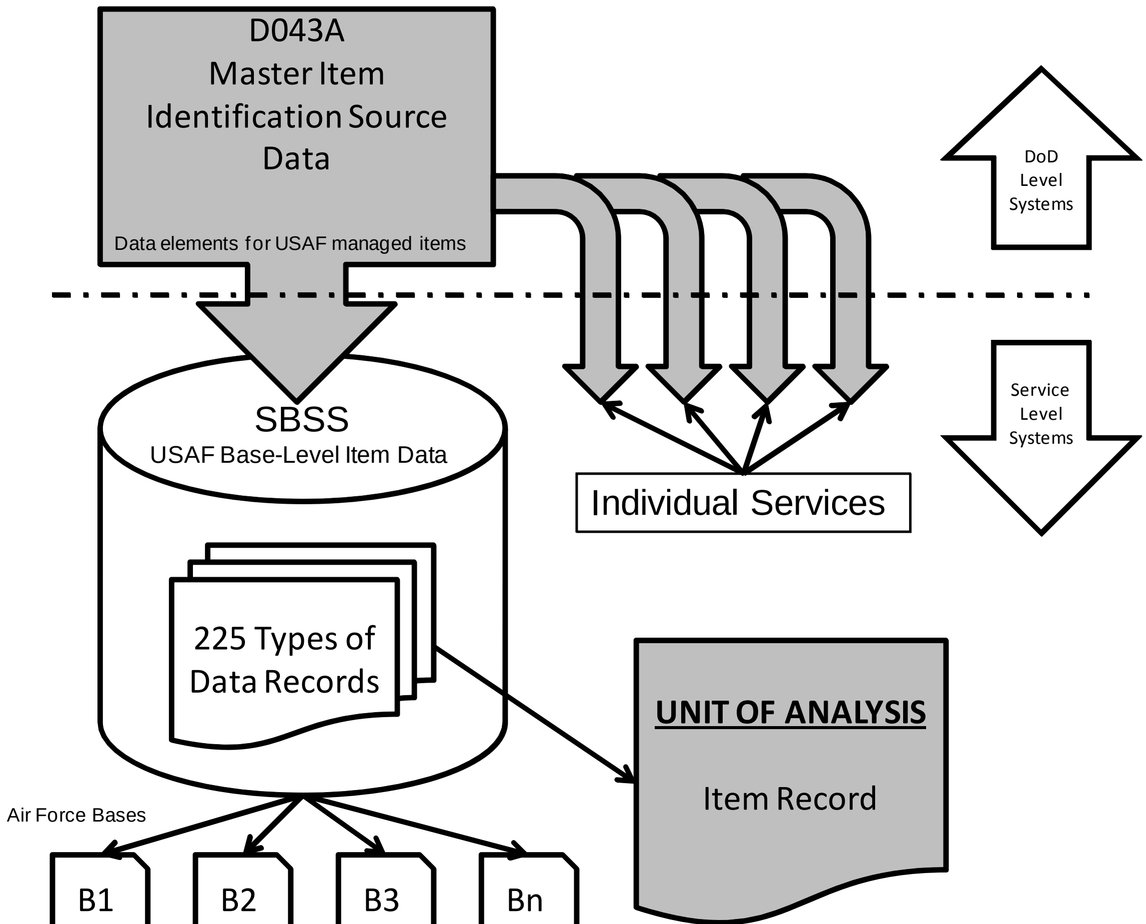


Figure 5 – Unit of Analysis

The item record and similar variants exist across several systems within the USAF. The data elements which comprise an item record are not all necessarily unique to the item record. Data elements may be duplicative and used among other types of data records in and among other legacy data systems. The authoritative source data for the data elements populating an item record in SBSS originate in D043A. D043A will be used as the control for comparison of the same data elements residing in item records extracted from SBSS. A visual representation of the verbiage used to describe this research is useful to identify the different data areas studied. Figure 6 distinguishes the structural breakdown of the data files utilized for analysis post-formatting.

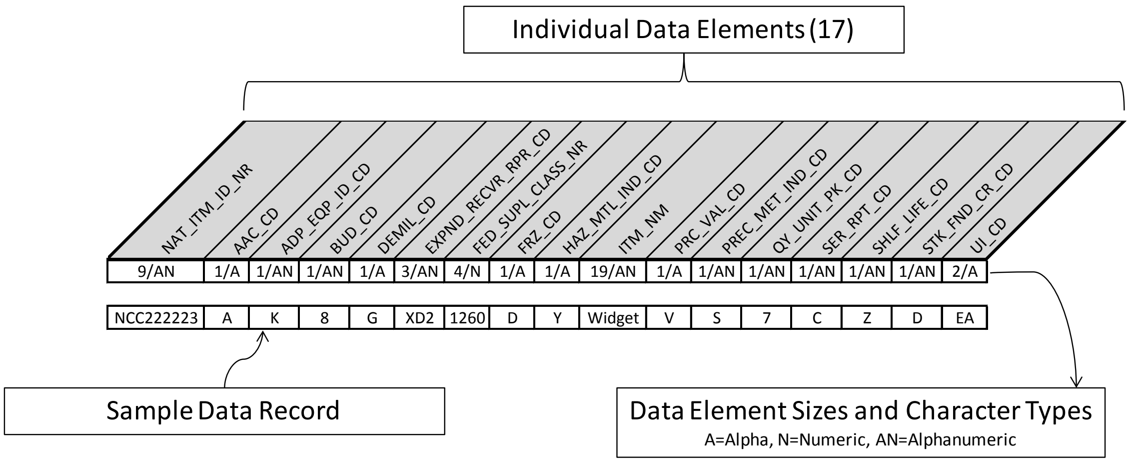


Figure 6 – Formatted Data File Structure

Research Design

The design selected for this research is a statistical analysis of the data collected. The end result is an attempt to determine which data elements may provide an indication of the quality of an item record. With this information, the Air Force can better allocate limited resources to focus data cleansing efforts on the areas where the greatest impact can be achieved prior to migration into ECSS. The data from D043A is the primary focus of the analysis, as this system will be treated as an authoritative source for populating data in ECSS.

A combination of regulatory guidance and advice from subject experts was used to determine the legacy systems and data selected for analysis, as well as how to appropriately analyze the data to derive significant and useful results. The regulatory guidance provided a baseline for answering portions of the investigative questions.

However, the constructive advice and input gleaned from experts who collectively share more than 4 decades of experience in the USAF Supply competency also proved to be quite valuable. This section provides a detailed discussion on the approach to segmenting and analyzing the data following an overview of the data sources and data collection. A straightforward methodology was designed to sequentially guide the research through each step of the data segmentation and analysis processes. A schematic of the designed research methodology is shown in Figure 7.

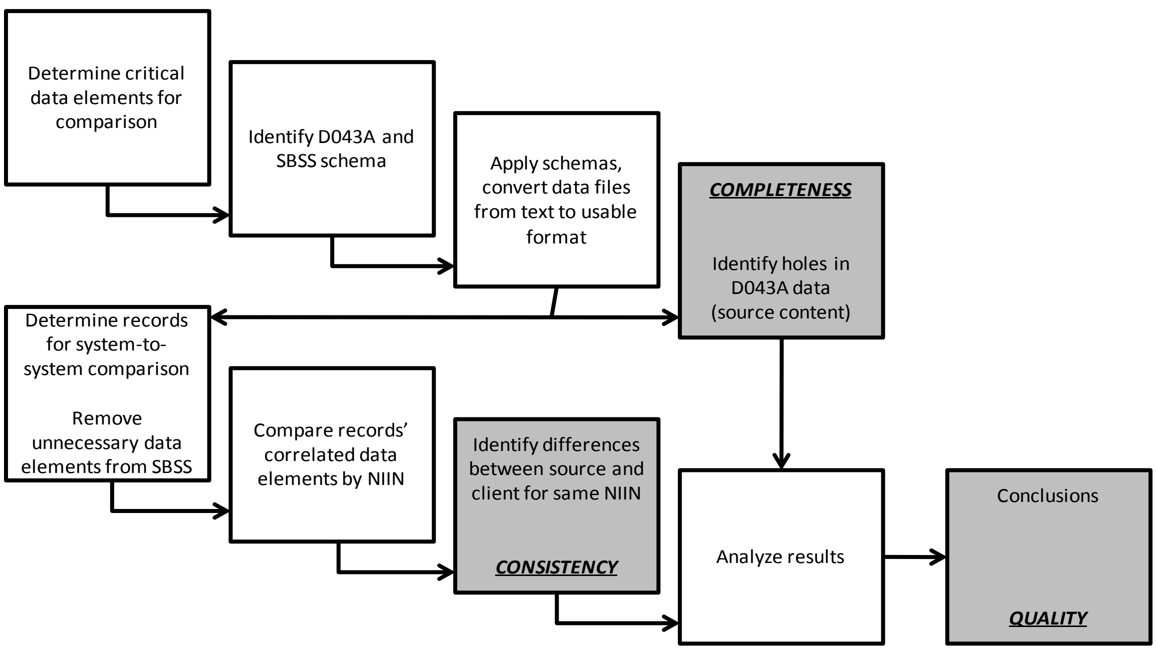


Figure 7 – Design of Experiment

Data Sources and Collection

The following legacy systems were selected for extracting the data necessary to complete the quantitative analysis conducted in this research:

- D043A Master Item Identification Database (MIIDB)
- D200A Standard Base Supply System (SBSS)

D043A enables web-based access to the data originating in D043, Item Management Control System (IMCS). As such, D043A provides access to data which serves as the authoritative source data for comparison and analysis in this research. D043 is the central repository of Federal and USAF logistics data for Air Force-used items of supply. It feeds several downstream service-level legacy systems including SBSS (AFMC, 2007). SBSS is the downstream system selected to provide data for comparison against the source data. It is reasonable to assume that data inconsistencies at the source create inconsistencies across all downstream systems.

Both data sets share the same baseline characteristics. They both represent a snapshot in time of all data records from the respective systems as of 31 December 2008. The D043A data extract includes only 17 data record elements, which correlate directly to SBSS, for USAF-specific items in the D043A database. The reason for the limited number of data elements is explained later in more detail. This data set was provided by the 401st Supply Chain Management Squadron (401 SCMS) which resides functionally under the Global Logistics Support Center (GLSC).

All SBSS data is USAF-specific by default because it is an Air Force system. The data extract from SBSS includes all item record transactions from every base across the entire USAF for the month of December 2008, as of the last day of the month. This data set was extracted from the Air Force Supply Data Bank and provided by the Air Force Logistics Management Agency. Both data sets represent the entire population of Air Force-specific data for the respective system being compared for analysis. Because this research started with the entire population, sample sizes were not a consideration. The subsequent methodology resulted in tailored data representing only the relevant items for final analysis.

Data Analysis

Ultimately, the end result of this data analysis is to identify data elements which can be used as potential indicators of data quality. This analysis compares the item record data elements which are common in both D043A and SBSS. Several prerequisite steps were necessary to refine both data sets in order to enable final comparison of the specific data elements common within both systems. This section outlines the entire analysis process sequentially and in detail.

The data files referenced in the previous section were received in a text file format as a single, continuous string of text. Due to the size of the files, there was no organic capability to manipulate them in any way. Qbase™, located in Dayton, Ohio, was instrumental in filling this critical gap between the raw data and its final analysis. In addition to their extensive data management experience, Qbase™ employed two of their proprietary tools, Qbase Data Discovery™ and Qbase Data Transformer™, to convert the raw data into a usable format for the detailed statistical analysis. These tools are designed to rapidly uncover data condition, report data anomalies and provide a rich visualization environment where source data SMEs and data experts can interact to understand exactly what can and cannot be accomplished with a given data set (Judson and Kinney, 2009).

With raw data management addressed, the initial task was to determine which data elements would be selected for comparison across the systems. The SBSS data file and its schema were received first. This provided the basis to work backward and request only the necessary data elements from D043A. Some preliminary work was completed prior to requesting the D043A source data file. Because D043A is an item identification data system for the entire Federal Government and the DoD, it contains an enormous amount of data. By identifying the specific data elements required for comparison ahead of time, the source data file was somewhat tailored at the point it was generated. This action saved time while still meeting the data needs of this study.

Regulatory guidance provided a foundation for selecting the data elements for comparison in this study. AFMAN 23-110, Volume 2, Part 4, Chapter 7, Attachment 7A-2, para 7A2.1, lists all 102 data elements contained in a SBSS item record. In the absence of a similar item record structure in D043A, a dummy sample data file was requested to determine what data elements were available from the system. This sample resulted in 60 data elements available for initial comparison to the SBSS data elements. The entire list of the data elements from the initial D043A dummy sample is displayed in Appendix D.

Several of the data elements residing in an item record found in SBSS are Air Force-specific. These data elements are assigned and populated at the service-level. Subsequently, they would not be found in D043A. The same principle is true of some D043A data elements as there are data elements in use at the Federal level which are of no use to the Air Force. Using the schema supplied with the data file, the SBSS data elements were compared to the data elements available in D043A. This comparison identified 17 correlated data elements to be used in the final analysis. This list was also used to request the D043A source data file so it would include the data needed for this study. Table 3 lists the correlated data elements, and their definitions, identified from both systems.

Table 3 – Correlated Data Elements

D043A Data Element	Definition	SBSS Data Element
AAC_CD	Acquisition Advice Code	ACQUISITION_ADVICE_CODE
ADP_EQP_ID_CD	ADPE Flag/Code	ADPE_FLAG
BUD_CD	Budget Code	BUDGET_CODE
DEMIL_CD	Demilitarization Code	DEMILITARIZATION_CODE
EXPND_RECVR_RPR_CD	Expendability/Recoverability/Repairability Code	ERRCD
FRZ_CD	Freeze Code	FREEZE_CODE
FED_SUPL_CLASS_NR	Federal Supply Classification	FSC
HAZ_MTL_IND_CD	Hazardous Material Indicator Code	HAZARDOUS_MATERIAL_CODE
ITM_NM	Item Name	NOMENCLATURE
NAT_ITM_ID_NR	National Item Identification Number (NIIN)	NAT_ITM_ID_NR
PRC_VAL_CD	Price Validation Code	PRICE_VALIDATION_CODE
PREC_MET_IND_CD	Precious Metal Indicator Code	PRECIOUS_METALS_FLAG
QY_UNIT_PK_CD	Quantity Unit Pack Code	QTY_UNIT_PACK_CODE
SER_RPT_CD	Serialized Report Code	SERIALIZED_REPORT_CODE
SHLF_LIFE_CD	Shelf Life Code	SHELF_LIFE_CODE
STK_FND_CR_CD	Stock Fund Credit Code/Flag	STOCK_FUND_CREDIT_FLAG
UI_CD	Unit of Issue	UNIT_OF_ISSUE

The next tasks included importing the files and formatting the text per the respective file schemas, which were supplied by the originators of the data files. A file schema defines how many character spaces are required for each data element in a continuous, single line of text. It may also define the specific character spaces a data element fills within a data record, i.e., columns 1 through 4. Additionally, the schema defines what separates or delimits the data characters to identify a data element in a continuous string of text as well as what type of characters the data element should be comprised of, e.g. comma, pipe, or tab; and alpha, numeric, or a combination of alpha-numeric. This step was critical to the remainder of the data segregation. Properly applying the schemas to ensure precise separation of the data elements contained in both files was imperative for an accurate and valid comparison across the systems later in the analysis. Appendices E and F illustrate the complete file schemas used to format the D043A and SBSS files respectively.

It has already been mentioned that the data analyzed in this study represents the population of Air Force-specific items in D043A as well as every item record in SBSS for the specified time period. However, before moving on it is significant to note the actual amount of raw data extracted, formatted, sorted, and analyzed at the onset of this study. Table 4 provides the raw numbers for each data file prior to any manipulation.

Table 4 – Initial Raw Data

System	Data Elements	Lines of Data
D043A	18	341,743
SBSS	106	3,420,181
Total		**3,761,924**

With both data sets converted from their text formats and ready for further segregation, the first portion of analysis could be addressed. Completeness was previously defined as one of the six characteristics of data quality (Becker, 2009). According to the LTO, the D043A database will eventually be a primary feeder to help populate ECSS when it comes online. Interrogation of the aggregate D043A data provides valuable insight about the current state of the data residing in the system.

The regulatory guidance, for D043A and SBSS, was researched in-depth to define all valid and acceptable parameters for each data element analyzed. This range of potential data entries for each data element provided the boundaries necessary to determine the completeness of the data analyzed. The D043A file was analyzed as a whole, and then each of the 17 data elements (refer to Table 3 above) was individually analyzed. Descriptive statistics were provided to support conclusions regarding the completeness of the entire D043A data set as well as specifics for each of the individual data elements contained within it.

The process of determining the consistency of item record data elements between the two systems started with identifying which records would be used in the comparison. The National Item Identification Number (NIIN) is a unique nine character code assigned to each item of supply purchased, stocked, or distributed within the Federal Government. It is used as the common denominator for an item of supply (AFMAN 23-110, Vol 2, Part 2, Ch 3, Para 3A1.2). For this reason, the NIIN was used to identify the same item and its associated data elements within both systems for comparison.

Because D043A is the source database, there should be only a single instance for any NIIN. Conversely, the SBSS data is transactional and it spans the entire Air Force meaning the same NIIN may occur at multiple bases due to common use and/or mission. This situation creates two distinct cases for the analysis of consistency: consistency between the source system and the client (downstream system), and consistency between the client systems all fed by the same source system. Figure 8 depicts the two cases created. Analysis of case 1 will produce results for case 2 based on design however, for the purposes of this research; only case 1 is analyzed in detail.

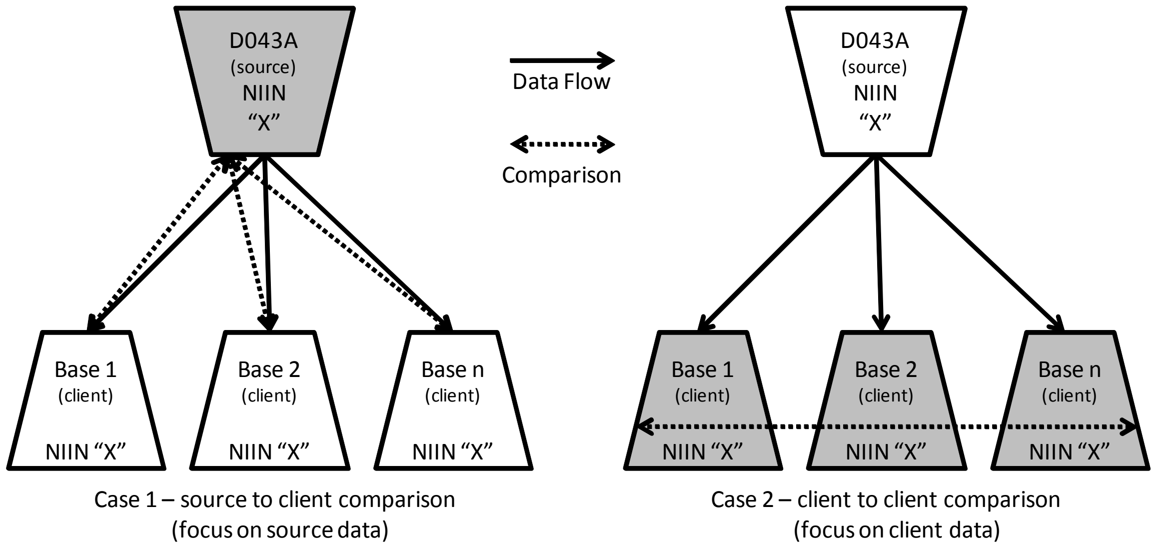


Figure 8 – NIIN Comparison Cases

For the actual record comparison between systems based on case 1, the data elements of the SBSS data file were pruned to match the data elements found in the D043A source data file. Of the 106 original data elements, all but 17 were removed from the SBSS data so the appropriate fields could be analyzed. In any instance where a NIIN did not reside in both data files, it was removed from the data set. This further paired the data, making it more manageable for the by-NIIN system to system record comparison based on the predetermined correlated data elements shown previously in Table 3.

The comparison identified any and all differences in the data elements between the source and client systems for the same NIIN. These results provided the foundation for analyzing the individual data elements (factors) which may be indicative of the overall quality of an item record. The in-depth analysis and results are provided in Chapter 4, followed by a presentation of the conclusions regarding the results of the analysis in Chapter 5.

Summary

This chapter outlined the important aspects of the quantitative and experimental aspects employed for this study. By defining which legacy data systems to sample, and more specifically what data from those systems to sample, the foundation was set to provide a manageable experiment. The investigative questions and expert input aided in legacy system selection, identification of an appropriate unit of analysis, and the specific data required for analysis. The data population was pruned using previously identified data elements and the NIIN. The analysis was completed using regulatory guidance to set parameters which would help establish a measure of data quality. By changing the unit of analysis and the focal data elements, this methodology should adapt easily to any system database experiment with a source-client relationship.

IV. Data Analysis and Results

Overview

This chapter contains the detailed results of the data analysis guided by the investigative questions and methodology outlined in Chapter 3. Before examining the results, the investigative questions are revisited and answers are provided. The completeness of the D043A data file is addressed first, followed by the results of the consistency comparison of item record data between D043A and SBSS. The results for both completeness and consistency include aggregate numbers as well as specific percentages for the individual data elements.

Investigative Questions and Answers

1. What are the valid data character entries for the analyzed data elements? Data requirements for D043A are governed by DoD 4100.39-M whereas SBSS is governed by AFMAN 23-110. The 17 data elements chosen for comparison and analysis were individually researched in both previously identified publications. The "Application Data" field was removed from the D043A source data file for the source-client comparison because the same field was not available in the SBSS data file. A matrix was developed using both sources to identify all possible entries for a given element. This information was used to establish the boundaries for determining the completeness of the source data. It also set the foundation for an accurate comparative analysis with the client data. Definitions for each of the data elements analyzed and their array of potential valid entries are listed alphabetically in Appendix G.

2. What constitutes a complete record for the purpose of analysis? The portion of data analysis in this study relating to record completeness focuses solely on the D043A data. As previously stated, D043A will be a primary data feed for migration into ECSS. All of the data elements (18) contained in the data source file are used as a basis for determining the overall completeness of a record. A record is deemed incomplete if the analytical software determines a particular data element value to be invalid in some way, e.g., null value or empty (when not valid), and/or an improper format per the schema. Complete records have all associated data elements populated (value present where/when required) and valid (properly formatted).

3. What constitutes a consistent record for the purpose of analysis? The consistency portion of the analysis includes both the source data (D043A) and the client data (SBSS), and uses the NIIN as a basis for comparing data elements across the systems. Consistent records will be identical to one another whereas inconsistent records will have dissimilar data contained within one or more of the correlated data elements. It is important to note here, although a source data record may have been identified as incomplete it can be identified as consistent. There are data elements where a null entry (empty field) is valid. These cases are addressed where and when necessary.

4. What constitutes a quality record for migration into the ECSS database? The importance of quality data regarding successful ERP system implementations cannot be overstated. The ECSS is a critical cog to aiding the successful transformation of the Air Force logistics enterprise. Having quality data is a paramount requirement to exploiting the full potential of ECSS. Furthermore, it is pivotal to achieving positive, effective results while implementing and developing an efficient ERP system. As such, for the

purposes of this research, quality records from the data analyzed are considered to be both complete and consistent.

However, there is one exception to this consideration on the basis of the analysis being limited to only two of the six quality data characteristics. In the event a source data record is deemed incomplete (null data elements only) and subsequently determined to be consistent with the correlated client data record, it will be treated as a quality record. This situation was also mentioned in the answer to investigative question number 3. In the absence of analysis focused on data element accuracy, the assumption is made that the null data element is justified and accurate.

Completeness

As previously stated, the analysis to determine the completeness of the data focused specifically on the D043A source data file, as this data is slated to be migrated into ECSS. The valid entry criteria listed in Appendix G were applied to the D043A data file to set boundaries for each of the individual data elements. Table 5 summarizes the amount of raw data analyzed followed by the results of the analysis in Figures 9 - 11.

Table 5 – Raw Data for Completeness Analysis

System	Data Elements	Lines of Data	Unique NIINs	Total Data Elements
D043A	18	341,743	341,743	6,151,374

Figure 9 represents the aggregate amount of invalid entries for each of the 18 individual data elements. It is important to note that null (empty) entries are generally treated as invalid. However, there are some circumstances where a null entry is valid, i.e.

freeze code. This anomaly was accounted for in all analyses. A comprehensive table including all raw numbers and individual percentages is available in Appendix H.

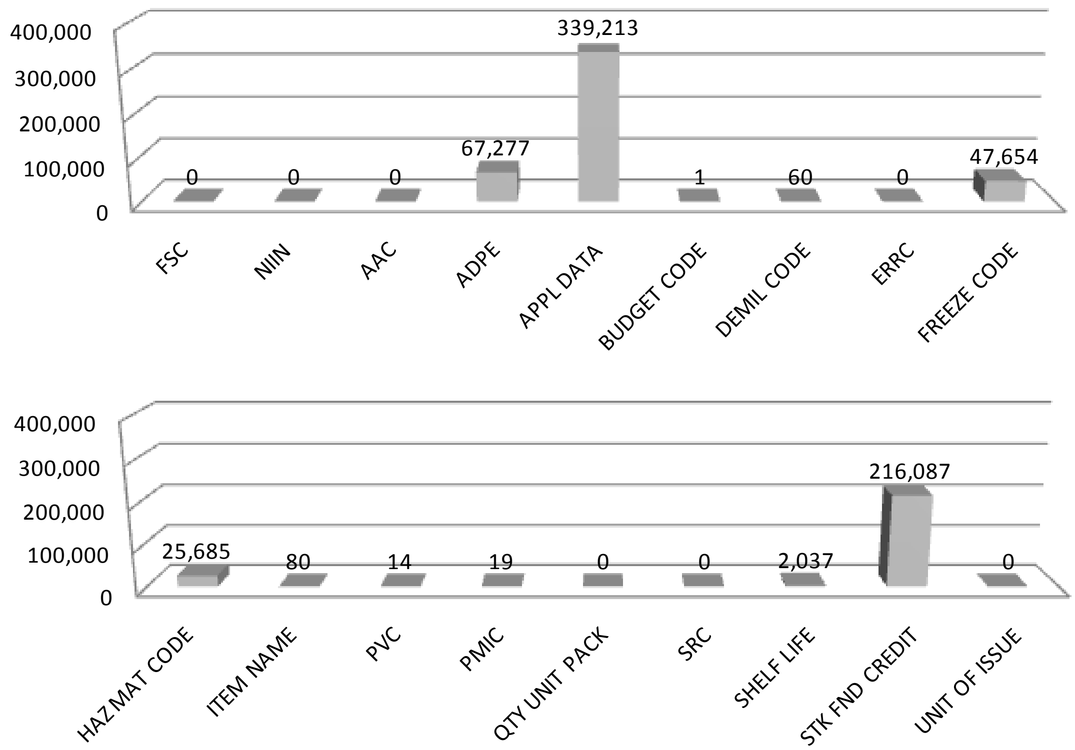


Figure 9 – Invalid Entries for Individual Data Elements

The aggregate numbers displayed in Figure 9 are translated into percentages in Figures 10 – 15 below. All depictions are in terms of data elements versus item records as the total number of discrepancies exceeded the number of records analyzed by moer than two to one. Figure 10 displays the amount of all invalid data elements compared to valid data elements, as a percentage of the total amount of data elements analyzed.

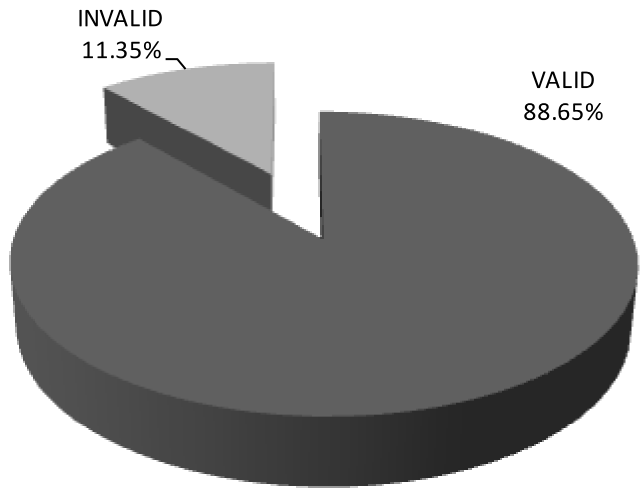


Figure 10 – Total Invalid Data Elements

Figure 11 displays all invalid entries within the data elements as a percentage of the total invalid entries. The data elements which accounted for less than 1% of the total invalid entries are collectively represented under the heading “OTHERS”. This representation highlights the largest areas of concern regarding invalid data entries.

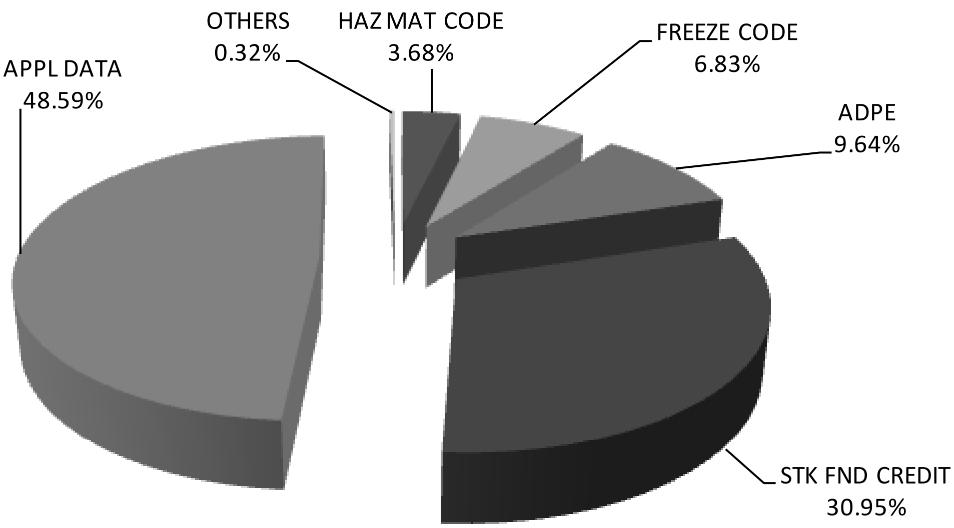


Figure 11 – Percentage of Invalid Entries

The Application Data (APPL DATA) element results accounted for almost half of the invalid entries. This data element is used to describe an item of supply for a specific system or platform. Every invalid entry was actually a null (empty) value so this data element was removed, and the statistics were recalculated. The conclusions in Chapter 5 provide a more in-depth explanation for this. Figure 12 displays the amount of all invalid data elements compared to valid data elements, as a percentage of the total amount of data elements analyzed, excluding the Application Data element. Excluding the Application data also affected the percentages of the individual data elements with regard to the total invalid entries. These statistics were also recalculated and are shown in Figure 13, which displays all invalid entries within the data elements as a percentage of the total invalid entries. The data elements which accounted for less than 1% of the total invalid entries are collectively represented under the heading “OTHERS”.

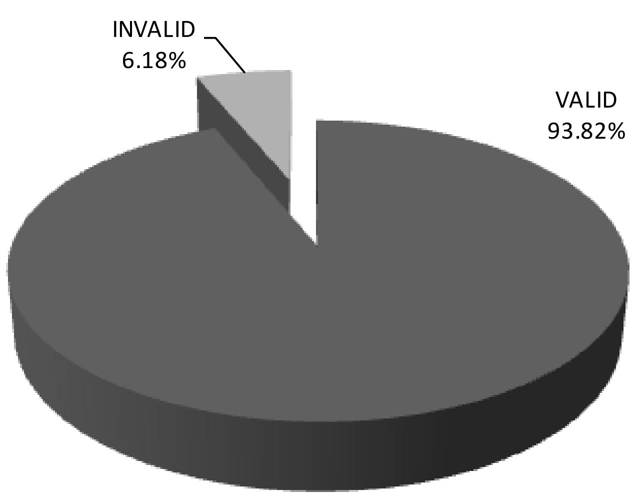


Figure 12 - Total Invalid Data Elements (excluding APPL DATA)

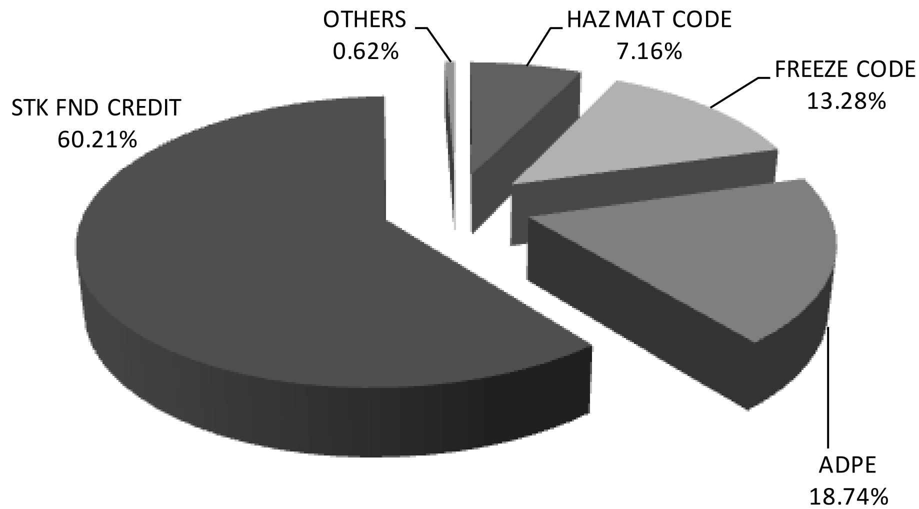


Figure 13 - Percentage of Invalid Entries (excluding APPL DATA)

After removing the Application Data (APPL DATA) element and recalculating the results, the Stock Fund Credit Flag (STK FND CREDIT) element accounted for more than half of the invalid entries. This data element is used to determine whether credit will be allowed for turning in an item of supply. Every invalid entry was actually a null (empty) value so this data element was removed, and the statistics were recalculated. The conclusions in Chapter 5 provide a more in-depth explanation regarding the removal of the Stock Fund Credit Flag data element. Figure 14 displays the amount of all invalid data elements compared to valid data elements, as a percentage of the total amount of data elements analyzed, excluding both the Application Data and Stock Fund Credit Flag elements. The effects on the individual percentages are shown in Figure 15. Following the same format, the data elements which accounted for less than 1% of the total invalid entries are collectively represented under the heading “OTHERS”.

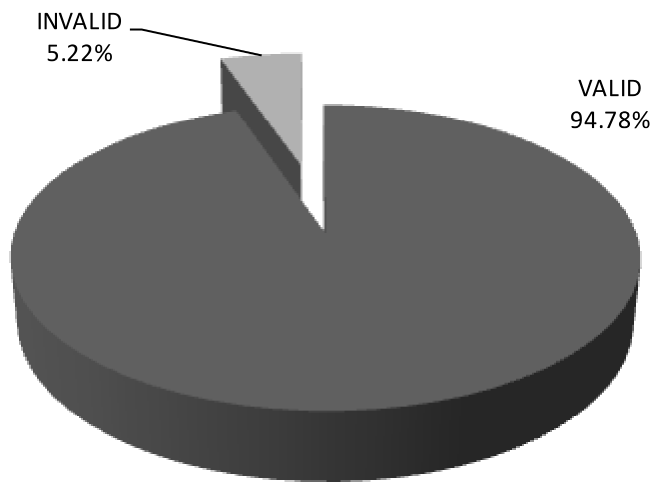


Figure 14 - Total Invalid Data Elements (excluding APPL DATA and STK FUND CREDIT)

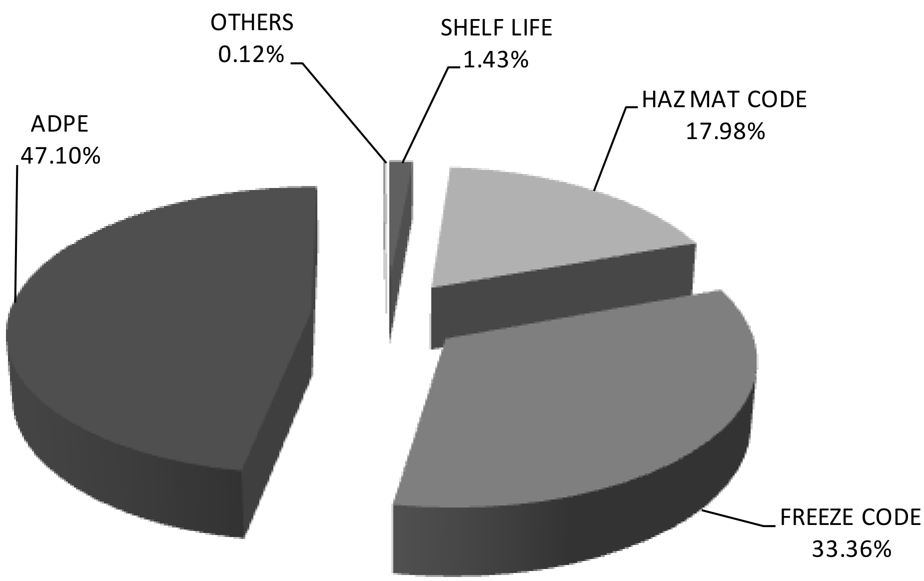


Figure 15 - Percentage of Invalid Entries (excluding APPL DATA and STK FUND CREDIT)

The depiction in Figure 15 highlights the data elements with invalid entries and their percentages among the total invalid entries. After removing both the Application Data and the Stock Fund Credit Flag data elements the ADPE data element becomes the top driver among all invalid data elements, accounting for 47.10% of all invalid data entries. The Freeze Code and HAZMAT Code are the second and third highest invalid data drivers at 33.36% and 17.98% respectively.

Consistency

The analysis to determine the consistency of the data utilized tailored data from both the D043A file and the SBSS file. It was first necessary to determine matching NIINs contained in both data files and exclude all others from comparison. Once the matching NIINs were identified, it was necessary to exclude all unrelated data elements from the comparison using Table 3 (found in Chapter 3) as a guide. As with the analysis for completeness, the valid entry criteria listed in Appendix G were applied to both data files to set boundaries for each of the individual data elements. Table 6 summarizes the amount of raw data analyzed followed by the results of the analysis in Figures 16 - 20.

Table 6 – Raw Data for Consistency Analysis

System	Data Elements	Lines of Data	Unique NIINs	Total Data Elements
D043A	17	126,833	341,743	2,156,161
SBSS	17	811,525	126,833	13,795,925
		938,358		15,952,086

Figure 16 represents the aggregate amount of inconsistencies (mismatches) for each of the 17 individual data elements between the D043A and SBSS files. For this portion of the study, the Application Data element was excluded as it did not exist in the

SBSS data file. A comprehensive table including all raw numbers and individual percentages for the data elements is available in Appendix I.

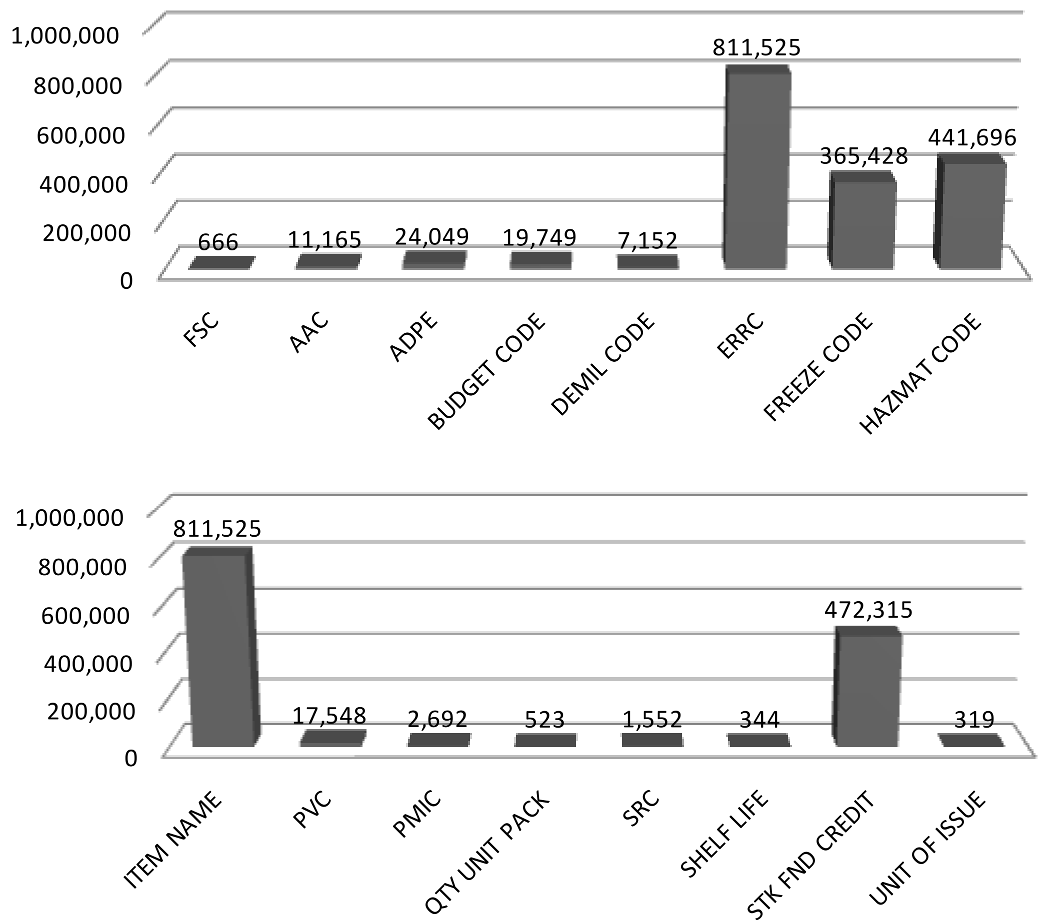


Figure 16 – Inconsistencies for Individual Data Elements

Figure 17 displays the amount of all inconsistent data elements compared to the consistent data elements, as a percentage of the total amount of data elements analyzed. Breakouts of the individual data elements with inconsistencies are presented as well to identify more specific areas of potential concern.

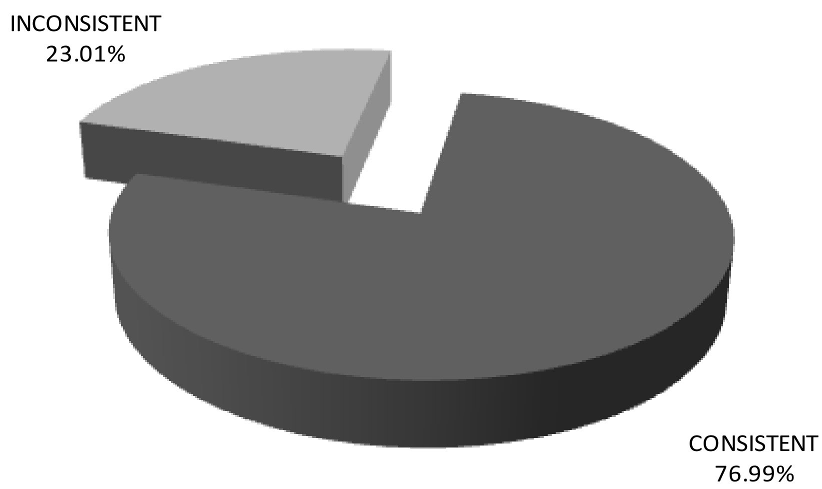


Figure 17 - Total Inconsistent Data Elements

Figure 18 displays all inconsistent data elements as a percentage of only the total inconsistencies. The inconsistent data elements which accounted for less than 1% of the total inconsistencies are collectively represented under the heading “OTHERS”. This representation highlights the largest areas of concern regarding inconsistent elements.

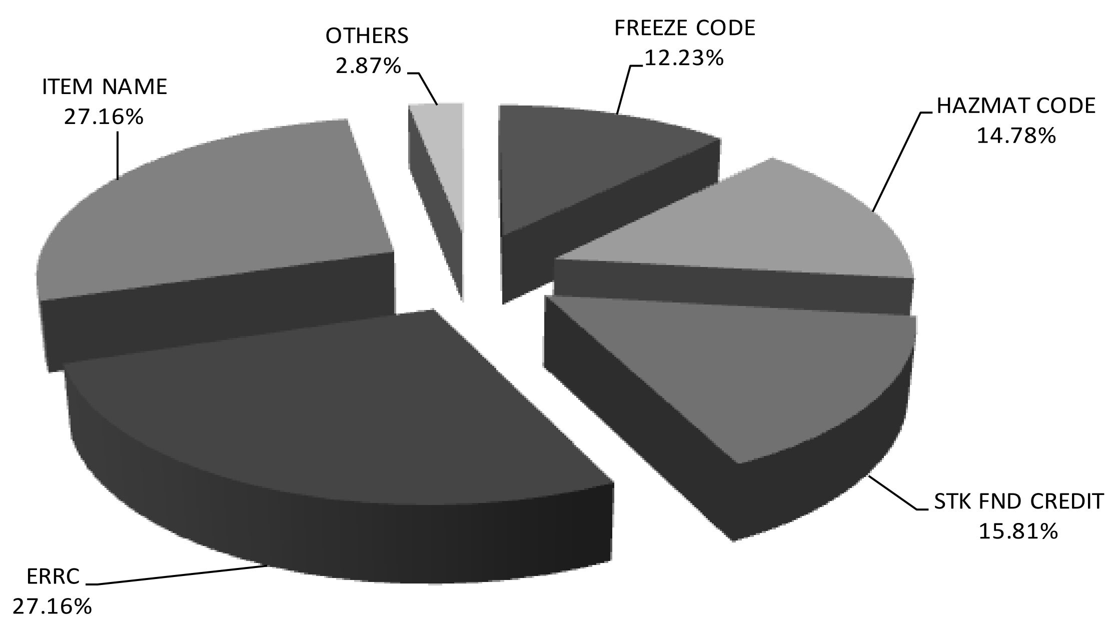


Figure 18 – Percentage of Inconsistent Data Elements

The Expendability, Recoverability, Reparability, Cost Designator (ERRC) Code and Item Name/Nomenclature data elements collectively accounted for almost half of the inconsistent data elements. It was discovered after analysis that the ERRC data element is coded differently between D043A and SBSS which led to a 100% mismatch between the data files. Also, the Item Name/Nomenclature data element is variable by definition as shown in Appendix G. This also represented a 100% mismatch between the data files. To provide a better level of fidelity, these data elements were removed. The statistics were recalculated excluding these two data elements and the results are shown in Figures 19 and 20 below. In Figure 20, following an already established format, the inconsistent data elements which accounted for less than 1% of the total inconsistencies are collectively represented under the heading "OTHERS".

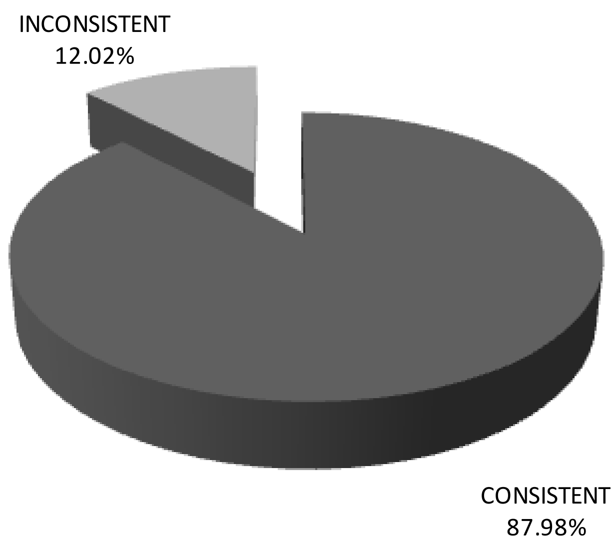


Figure 19 - Total Inconsistent Data Elements
(excluding ERRC and Item Name/Nomenclature)

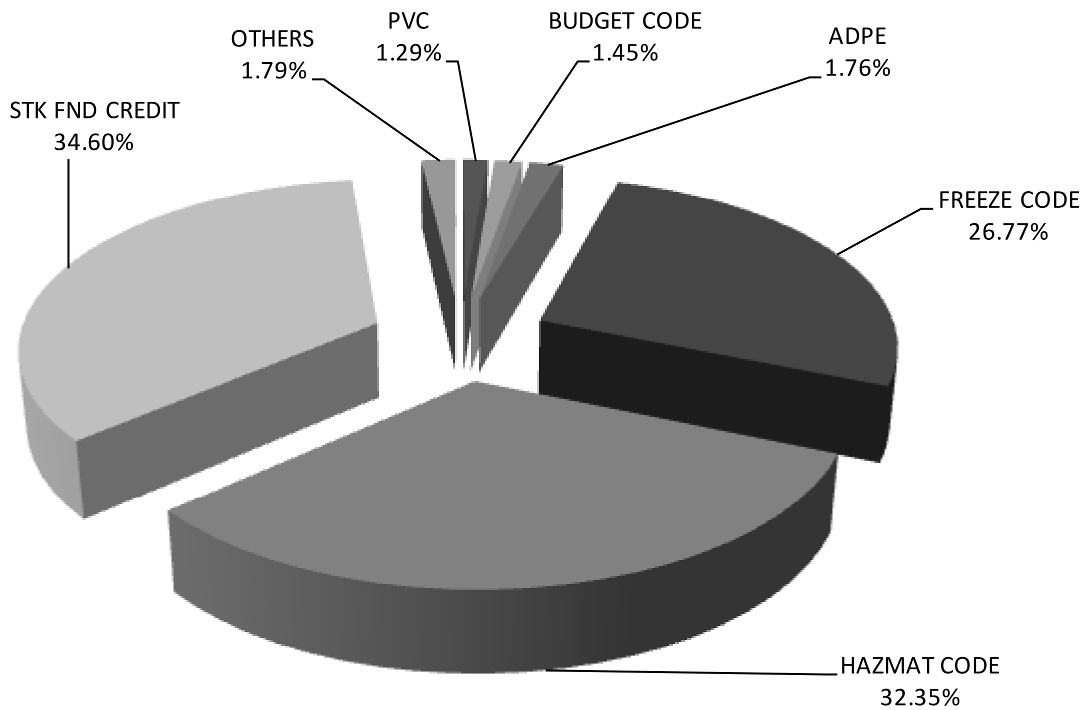


Figure 20 – Percentage of Inconsistent Data Elements (excluding ERRC and Item Name/Nomenclature)

The depiction in Figure 20 highlights all the data elements which have inconsistencies and their percentages among the total inconsistent data elements. After removing both the ERRC Code and the Item Name/Nomenclature data elements the Stock Fund Credit Flag data element becomes the top driver among all inconsistent data elements, accounting for 34.60% of all inconsistent data elements. The HAZMAT Code and Freeze Code are the second and third highest invalid data drivers at 32.35% and 26.77% respectively.

Summary

This chapter revisited and answered the investigative questions of this research. These answers provided the framework for the subsequent data analysis. An extensive analysis of the data was completed and the results were provided. Additionally, some data elements were excluded and alternate scenarios were explored using assumptions shaped by the initial results. These additional results were also presented which provided more fidelity for developing the conclusions about data completeness and consistency discussed in Chapter 5.

V. Conclusion

Overview

This final chapter serves to sum up the entirety of this study. Following the results of the analysis from the previous chapter, the research questions are revisited and answered. Additionally, the researcher's conclusions and recommendations are stated. Some lessons observed during the course of this work are provided for the benefit of anyone continuing on with similar research. The chapter concludes with a brief discussion about the assumptions and limitations of the study, and areas for future research.

Research Questions and Answers

1. How complete are item records? Based on the results of the analysis for completeness (listed in Appendix H), the answer to this question is dependent upon other factors excluded (not purposely) from this research. The total number of invalid entries more than twice exceeds the total amount of records analyzed. This means any given record could have at least one or multiple invalid entries. Without analysis to correlate each invalid entry to a specific NIIN, it is not possible to determine a concrete level of completeness. For that reason, a range encompassing potential completeness was developed based on the results.

The Application Data element contained the highest percentage of the total invalid entries, 339,213 or 99.26%. In this case, all invalid application data entries were actually null values meaning the element was empty. The assumption made by the researcher was that the Application Data items were likely common-use items across several systems or platforms, and therefore would not have unique data for this field. Using this assumption, the Application Data element was removed and the statistics were recalculated.

After recalculation excluding the Application Data, the numbers showed that the Stock Fund Credit Flag data element contained the highest percentage of the total invalid entries, 216,087 or 63.23%. Upon further investigation, it was discovered that all invalid entries were actually null values meaning the element was empty. The guidance regarding this data element is not explicit, however, it is the belief of the researcher that this data element is dependent on an item's ERRC code. According to the regulatory guidance, there are two possible values for this data element. One entry allows credit for an item, while the other one does not allow credit. Based on the fact some items can be consumed in use and are expendable, the assumption was made by the researcher that a null (empty) value is also valid. For this reason, the Stock Fund Credit Flag data element was also removed and a recalculation of the statistics was completed.

This final set of results was used to calculate a range regarding the completeness of the D043A data file. After excluding the two previously identified data elements, the Automated Data Processing Equipment (ADPE) data element contained the most invalid entries, with 67,277 or 19.69% of the total records analyzed. These numbers imply that 80.31% of the entries for this data element were valid. The sum of the remaining invalid data element percentages totals 22.10%. Therefore, assuming data accuracy and invalid entry independence, the valid range for potentially complete records is from 58.21% - 80.31%. In terms of aggregate numbers, of the 341,743 total records analyzed, we can reasonably expect at least 198,931, and no more than 274,466 to be complete.

2. How consistent are item records? The comprehensive results of the analysis for consistency are listed in Appendix I. The total number of inconsistencies exceeded the number of records analyzed more than threefold. This presents a similar issue addressed in attempting to identify completeness. Any given record could have at least one or multiple inconsistencies.

The initial results identified two data elements which were 100% inconsistent between the systems: Item Name/Nomenclature and the Expendability, Recoverability, Reparability, Cost Designator (ERRC) Code. The regulatory guidance, for both D043A and SBSS, regarding the Item Name/Nomenclature is contradictory. Using a combination of both regulations, a worst-case parameter was developed for this data element; 19 – 32 characters and alphanumeric. A high mismatch percentage was expected as almost any value is valid from a computing perspective.

The ERRC code presented a different problem. The reason for the 100% mismatch was discovered as the data was being processed, but at a point too late to fix. While the regulatory guidance for both systems is congruent, the proverbial "fine print" is critical to linking the ERRC code between the systems. The ERRC code is a 3-character alphanumeric code. In the interest of physical space in the data system, D043A utilizes an ERRC code designator, a single alphabetic character which correlates directly to the ERRC code in SBSS. The specific characters are listed in Appendix G.

Because these two data elements were quite likely skewing the results, they were removed from the data set and the statistics were recalculated. The intent in determining consistency was to follow a similar format as was used to determine completeness, i.e. develop a potential range regarding consistency. However, even with the 100%

mismatches removed, the number of inconsistencies exceeded the total number of records compared. As with completeness, without a specific correlation of inconsistencies by record between the two systems, it is difficult to determine a concrete level of consistency. Assuming the errors were independent of each other, meaning there was at least one error per record, this implies there is limited consistency in the data between the systems.

3. Where should resources be allocated to address data cleansing/correction? According to the results of this study, there are three data elements comprising the bulk of invalid entries for completeness: Automated Data Processing Equipment (ADPE) Code, Freeze Code, and Hazardous Materials Code. The majority of the inconsistencies also focus on three data elements: Stock Fund Credit Flag, Hazardous Materials Code, and Freeze Code. In terms of allocating resources to data cleansing, the results show the Hazardous Materials Code and the Freeze Code are points of concern in both sets of analysis. For this reason, they should be the first priority. The ADPE Code would be the next data element for focus. While the Stock Fund Credit Flag was removed from the completeness analysis, it represented the highest percentage of inconsistency. This also requires some resolution.

4. What are the potential implications of these results? The quality of the data is a critical key to any successful ERP system implementation. This fact is addressed at length in Chapter 2 of this study. Data quality enhances system performance, builds trust in the system among users, and provides leadership with accurate information for better decision making. While this study focuses only on two quality data characteristics and two systems, one of which will provide data for migration into ECSS, it highlights some

of the potential flaws with existing data. Assuming the data in this study are somewhat representative of the several hundred systems to be consolidated by ECSS, data quality viewed even in a broad perspective is questionable at best. Using this data to populate the new ERP system without first addressing its overall quality has the potential to impede the path to a successfully implementation of ECSS.

Additional Findings

The format of this study was fairly well scoped to keep the project focused and manageable. However, the nature of its design and the subsequent analysis of the data for consistency derived other results, which although peripheral to this study, are significant in terms of the costs associated with some of the items studied. Table 7 shows the aggregate number of mismatches by data element as well as the sum of the unit prices for the items whose records mismatched for that specific data element.

Table 7 –Cost of Mismatched Items

Data Element	Mismatches	Total Cost of Items with Mismatches
ERRC	811,525	$26,375,814,635.81
NOMENTCLATURE	811,525	$26,375,814,635.81
FREEZE CODE	365,428	$14,841,002,205.78
STOCK FUND CREDIT FLAG	472,315	$13,623,549,950.32
HAZMAT CODE	441,696	$12,659,083,969.38
ADPE CODE	24,049	$632,672,959.86
PRICE VALIDATION CODE	17,548	$561,654,664.42
AAC	11,165	$353,730,040.61
BUDGET CODE	19,749	$241,889,620.16
DEMIL CODE	7,152	$203,324,691.33
SERIALIZED REPORT CODE	1,552	$133,091,897.50
PRECIOUS METALS INDICATOR CODE	2,692	$78,277,838.13
UNIT OF ISSUE	319	$69,159,245.09
SHELF LIFE CODE	344	$50,493,101.07
QTY UNIT PACK CODE	523	$47,811,174.98
FSC	666	$3,707,013.86

These costs are not shown to imply lost capital. However, it would seem intuitive that Table 7 highlights yet another reason to address the issue of data quality. Not only is the cost of the items which are managed with bad data staggering, but it further opens the possibility for associating an actual cost to that bad data.

Recommendations

The point of this study was to identify specific areas within system data to focus cleansing efforts. The results of the data analysis highlighted the data elements with the highest percentages of invalid and/or inconsistent entries. The specific data elements used for this study were selected because they were consistent between the source and client systems. It is possible that some of these elements, and several others not analyzed, may not be migrated into the new system. Therefore, identification of the elements being carried forward to ECSS and eliminating those which are not, would serve to focus data cleansing efforts.

It would also be beneficial to apply the methodology of this study to other systems which will be consumed by ECSS. The importance of data quality regarding the implementation of ECSS is not limited to the systems and data studied in this research. Extending this type of study to other systems and comparing the results with those presented in this research would provide a more accurate representation of the quality of data in our existing legacy systems.

The ultimate recommendation as a result of this research is for the USAF to address data quality in existing legacy systems before migrating any of the data into ECSS. While the research presented in this work may have areas for improvement, it has served to identify a significant problem with existing data. If the intent of the USAF ERP implementation effort is to leverage industry best practices and lessons learned, then the literature review alone should provide the justification for pre-implementation data cleansing.

Due to the large amount of data element inaccuracies, it was impossible to use an item record as a basis for comparison when defining results. All results were broken out in terms of the invalid and/or inconsistent values. Intuitively, this researcher believes mapping individual data elements as opposed to mapping data records would much better serve the data efforts regarding the implementation of ECSS. First, specifically identify which data elements are needed in the new system. Answer this question: "what data do we need?" versus "what data do we have?" Then apply this methodology to those data elements in the existing systems slated to populate ECSS. Due to the fact that our existing legacy systems are several decades old, the some of the data in them may not be needed in the future state of the logistics enterprise. It is quite possible in this instance with cutting edge technology in our grip, less may be more.

Lessons Observed

This study revealed several valuable insights for the researcher. Data, as a general and broad topic, is universally important with regard to ERP system implementations. However, when drilling down to a specific area of concern or system to study, the challenges grow considerably. There is no shortage of experts on the individual systems or the data residing in those systems. Access to the explicit knowledge via regulatory guidance is virtually unlimited, though the regulations are exceedingly large and not easily navigated. Access to the tacit knowledge of system experts and the data residing in the systems is more limited. There are obvious security concerns with access to the data, but there also seemed to be a certain amount of a "pride in ownership" attitude regarding the possibility of having data issues identified. It was difficult and time consuming to finally get connected to the individuals who provided the data for analysis.

The amount of data residing in DoD systems is enormous. Despite the importance of these data, there is an obvious lack of interoperability. This hinders informed decisions and compounds inefficiencies across the enterprise. Transformation across a joint environment is the basis for many initiatives within the DoD today. It is difficult to be effective or efficient without a standard. This research further highlighted an already identified need for a common data standard.

Data analysis is an exceptionally rotund elephant…that regenerates. Even a sequential bite at a time seems counter-productive. Despite following a specific research design and methodology, it was difficult to remain focused. As each step of the analysis was completed, collateral damage followed in the form of unexpected findings and/or other potential concerns with the data, causing both doubt and hesitation. It seemed intuitive to chase these other rabbits, however, with limited time and resources these items were left to future research.

Assumptions and Limitations

When fully implemented, ECSS will have consolidated several hundred legacy systems. The research and analysis presented in this paper focuses on only two systems as points of comparison. Despite using the entire population residing in both systems for analysis, it is important to note this represents only a small fraction of the data currently residing in all affected Air Force legacy systems. Furthermore, the data used for analysis represents a specific point in time. It is reasonable to assume the data can, and may have, changed since the analysis occurred. Additionally, the unit of analysis is merely one type of record generated for use in these systems and as such, contains only a fraction of the potential data elements which exist across all systems for items in the Air Force inventory.

In the absence of concrete guidance regarding valid data characters allowable for specific data elements, personal judgment was used to make a decision regarding how to best frame the analysis of those specific elements. Regulatory guidance was used to the extent available. Coupled with the existing, yet limited information available within the data files, informed decisions were made concerning what constitutes valid entry data in the D043A data for the Freeze Code and the Stock Fund Credit Flag data elements. These assumptions are captured in Appendix G for the respective data elements.

Six characteristics proposed to define quality data were identified in Chapter 2 of this research. The methodology and data analysis of this study focused on only two of those characteristics, specifically completeness and consistency. It is conceivable that more detailed analysis on a smaller set of similar data utilizing all six characteristics has the potential to produce different results. In terms of analyzing completeness and consistency, the assumption was made that the existing data was accurate.

The intent of this work is to identify the factors which can assist in focusing limited resources on identifying and correcting the most inaccurate data within the studied systems. It is assumed that the methodology developed and implemented for this research can be applied to other legacy systems to identify areas for focus within them. Because the results of this study represent only a single type of data record across two systems, they are not assumed to be representative of all legacy systems. Only research on those specific systems would provide viable results.

Future Research

While this research serves to address some important questions surrounding the issue of data quality, it by no means answers all of them. First and foremost, the end result of this study was determining where to best allocate limited resources to effectively focus D043 data cleansing efforts prior to migrating data from the legacy system to ECSS. Data cleansing is no small task, especially concerning a project the size of ECSS. Developing an effective, empirically-based cleansing plan to address “dirty” data prior to migration would be a logical corollary to this work.

In terms of the proposed data terminology used throughout this research, a more rigorous study including all six characteristics of data quality may be in order. The assumptions and limitations of this study highlight some areas which require more significant probing. The analysis of the completeness of the D043A source data file was limited to 18 data elements. However, an analysis of completeness inclusive of all data elements found in the D043A database would likely provide its own unique results regarding this data quality attribute.

In addition to a deeper study of the completeness of existing data, a study focused on the accuracy of the existing data, while very labor-intensive, would be significant. For the purposes of this research, the assumption was made that the existing data used for analysis was accurate. The assumption of accuracy is loaded with substantial risk. A detailed, by-item study of the individual data elements contained in the source database would reveal the extent of inaccurate data and lay the foundation to significantly reduce any risks associated with “dirty” data. Furthermore, this idea of accuracy can be extended beyond the computer systems to the physical items on a shelf, i.e. the inventory data contained in the computer matching items held in inventory.

The data analysis for consistency highlighted a system issue. Across the Air Force, all base-level data for cataloged items originates at the same source. This implies that for cataloged items of supply, all data for that item (with minor exceptions) should be the same at all bases. The results show there is inconsistency between bases for the same item. This would imply that there are connectivity issues between the source (D043A) and the individual clients (bases). A strict comparison of cataloged items using base-level data would assess the magnitude of this issue and identify potential action items to address before ECSS is brought on-line.

Despite universal agreement regarding the importance of data quality, the field is broad, diverse, and in the researcher’s opinion, under-explored. This study alone is only a small step to aid in closing the void regarding both how, and where, to address data quality issues prior to ERP implementation. These focus areas recommended for future research serve to potentially bridge more of those gaps discovered throughout the course of this study.

Appendix A – ERP Timeline Acronyms

EOQ:	Economic Order Quantity
ROP:	Reorder Point
MRP:	Material Requirements Planning/Manufacturing Resources Planning
MRP II:	Material Requirements Planning/Manufacturing Resources Planning
DRP:	Distribution Requirements Planning/Distribution Resources Planning
FAX:	Facsimile Transmission
EDI:	Electronic Data Interchange
JIT:	Just-in-Time
QR:	Quick Response
CPR:	Continuous Product Replenishment
ECR:	Efficient Consumer Response
TOC:	Theory of Constraints
VMI:	Vendor Managed Inventory
ARP:	Automatic Replenishment Programs
RF:	Radio Frequency Systems
MES:	Manufacturing Execution Systems
ERP:	Enterprise Resource Planning
APS:	Advanced Planning Systems
XDM:	Extended Decision Management
CPFR:	Collaborative Planning Forecasting and Replenishment
CRM:	Customer Relationship Management
RFID:	Radio Frequency Identification
ERP II:	Enterprise Resource Planning (more of a supply chain/external focus)
ECM:	Enterprise Commerce Management (same concept as ERP II)

Fawcett et al., (2007)

Appendix B – SCOR Model Business Process Definitions

Plan – includes strategic and tactical planning, and accountability/reporting (overall management, administration, finance, accounting, and human resource management)

Source/Sell – from the supplier's point of view this is the customer order process, whereas from the buyer's point of view this is the purchasing/sourcing process

Make – involves the production, manufacturing, assembly, or service delivery process

Delivery/Return – both involve the logistics, warehousing, and transportation processes

Fawcett et al., (2007)

Appendix C – Quality Data Characteristic Definitions

Accuracy – correctness; degree to which the reported information value is in conformance with the true or accepted value

Consistency/Validity – degree of freedom from variation or contradiction; degree of satisfaction of constraints (including syntax/format/semantics)

Completeness/Brevity – degree to which values are present in the attributes that require them; degree to which values not needed for decision making are excluded

Timeliness – time/utility; degree to which specified data values are up to date

Pedigree/Lineage/Provenance (Authoritative) – history of data origin (also called lineage or provenance) and subsequent transformation

Precision/Certainty – exactness or confidence in value (vs. imprecise, uncertain, approximate, probabilistic, or fuzzy)

Becker, (2009)

Appendix D – D043A Data Elements (dummy sample)

Data Element ID	Attribute
AAC_CD	Acquisition Advice Code
ACQ_METH_CD	Acquisition Method Code
ACT_ITM_MGR_CD	Action Item Manager Code
ACTL_UNIT_PRC_AM	Actual Unit Price
ADP_EQP_ID_CD	Automated Data Processing Equipment Identification Code
AIR_FRC_ITM_MGR_CD	Air Force Item Manager Code
ALFT_ITM_CD	Airlift Item Code
AMSC_CD	Acquisition Method Suffix Code
APPL_DATA_TX	Application Data Transfer
BATCH_INS_NR	Batch Insurance Number
BATCH_UPD_NR	Batch Update Number
BUD_CD	Budget Code
CAT_ACTY_CD	Category Activity Code
CREATE_DT_TM	Create Date Time
CRIT_CD	Critical Code
DEMIL_CD	Demilitarization Code
DIPEC_CD	DIPEC Code
DIV_MGR_DESIG_CD	Division Manager Designator Code
DW_END_DT	DW End Date
DW_START_DT	DW Start Date
EFF_DT	Effective Date
ELEC_DSCHRG_CD	Electrostatic Discharge Code
EMC_CD	Equipment Management Code
EQP_SPCL_CD	Equipment Specialist Code
EXPND_RECVR_RPR_CD	ERRCD
FED_SUPL_CLASS_NR	Federal Supply Classification Number
FED_SUPL_GRP_NR	Federal Supply Group Number
FIIG_NR	Federal Item Identification Guide Number
FND_CD	Fund Code
FRZ_CD	Freeze Code
HAZ_MTL_IND_CD	Hazardous Material Indicator Code
I_S_IND_CD	Interchangeable & Substitute Code
ITM_MGR_DESIG_CD	Item Manager Designator Code
ITM_MGR_NM	Item Manager Name
ITM_MGR_OFF_SYM_TX	Item Manager Office Symbol
ITM_NM	Item Name
ITM_NM_NR	Item Name Number
JNT_MGT_CD	Joint Management Code
LL_CD	Lean Logistics Code (2-level maintenance flag)
MTL_MGT_AGG_CD	Material Management Aggregation Code (MMAC)
MUN_IND_CD	Munitions Indicator Code
NAT_ITM_ID_NR	National Item Identification Number (NIIN)
PRC_VAL_CD	Price Validation Code
PRC_VAL_DT	Price Validation Date
PRCUR_SRC_CD	Procurement Source Code
PREC_MET_IND_CD	Precious Metal Indicator Code
QY_UNIT_PK_CD	Quantity Unit Pack Code
RAD_CD	RAD Code
REF_PARTL_DES	Unknown
SCTY_CLASS_CD	Security Class Code
SER_RPT_CD	Serialized Report Code
SHLF_LIFE_CD	Shelf Life Code
SRC_SUPL_CD	Source Supply Code
STK_FND_CR_CD	Stock Fund Credit Code
SUPL_MGT_GRP_CD	Supply Management Grouping Code
TEL_NR	Telephone Number
TYP_ITM_ID_CD	Type Item Identification Code
UI_CD	Unit of Issue Code
UNIT_ISS_CNVER_RT	Unit of Issue Conversion Rate

Appendix E – D043A Schema

Data Element Name	Size	Type
A=alpha, N=numeric, A/N=combination		
FSC	4	N
NIIN	9	A/N
AAC	1	A
ADPE	1	A/N
APPL DATA	28	A/N
Budget Code	1	A/N
Demil Code	1	A
ERRC	3	A/N
Freeze Code	1	A
HAZ MAT Code	1	A
Item Name	19	A/N
Price Validation Code	1	A
Precious Metal Indicator code	1	A/N
Quantity Unit Pack Code	1	A/N
Serialized Report code	1	A/N
Shelf Life Code	1	A/N
Stock Fund Credit Flag	1	A/N
Unit of Issue	2	A

Appendix F – SBSS Schema

Data Element Name	Size	Type		Data Element Name	Size	Type
A=alpha, B= binary, N=numeric, A/N=combination						
SRAN	4	A/N		LOCAL_ERRCD_FLAG	1	A
STOCK_NUMBER	15	A/N		LOCAL_PURCHASE_FLAG	1	A
FSC	4	N		LOT_SIZE_FLAG	1	
NIIN	9	A/N		MANAGER_DESIGNATOR_CODE	3	
MMC	2	A		MAX_LEVEL_FLAG	1	A
ACQUISITION_ADVICE_CODE	1	A		MIN_LEVEL_FLAG	1	A
ADPE_FLAG	1	A/N		MISSION_CHANGE_GAIN_FLAG	1	A
AF_RAMPS_REPORT_CODE	1	A/N		MISSION_CHANGE_LOSS_FLAG	1	A
AFTO_FORM_95_CODE	1	A		MISSION_IMPACT_CODE	1	N
AIRLIFT_INVESTMENT_FLAG	1			MSK_RCD_FLAG	1	A/N
APPLICATION_CODE	2	A/N		MULTIPLE_DIFM_FLAG	1	A/N
BASE_CLOSURE_FLAG	1	B		NAT_MTR_FRT_CLASSTN	6	N
BENCH_STOCK_RCD_FLAG	1	A/N		NBR_DMNDS_007SC	3	
BOQ_CONSUMPTION_RCD_FLAG	1	A/N		NBR_OF_DMDS_CURRENT	2	
BUDGET_CODE	1	A/N		NBR_OF_DMDS_PAST_6_MONTHS	2	
CONTROLLED_ITEM_CODE	1	A/N		NBR_OF_DMDS_PAST_7_12_MTHS	2	
CSMS_REPORT_FLAG	1	A		NOMENCLATURE	19	A/N
CUMLTV_DEMAND_QTY	7			OST_OVERRIDE	3	
CUMLTV_DEMAND_QTY_SQ	15			OVERFLOW_ADJUNCT_RCD_FLAG	1	A/N
CUMLTV_RECURRING_DEMANDS	7	N		PRECIOUS_METALS_FLAG	1	A/N
CURRENCY_RCD_FLAG	1	A/N		PRICE_VALIDATION_CODE	1	A
D028_LEVEL_FLAG	1	N		PROBLEM_ITEM_FLAG	1	A/N
DATE_OF_FIRST_DEMAND	4	N		QTY_UNIT_PACK_CODE	1	A/N
DATE_OF_LAST_DEMAND	4	N		RBL_FLAG	1	
DATE_OF_LAST_INVENTORY	4	A/N		RELATIONSHIP_CODE	1	A
DATE_OF_LAST_RELEVELING	4	N		REQUIREMENTS_COMP_FLAG	1	
DATE_OF_LAST_SNUD_UPDATE	4	N		REX_CODE	1	A/N
DATE_OF_LAST_TRANSACTION	4	N		RID	3	A/N
DATE_OF_LAST_TRANSP_UPDATE	4	N		SAMPLE_INV_LOT_FLAG	1	
DATE_SPC_ASSIGNED	4	N		SERIALIZED_REPORT_CODE	1	A/N
DEMAND_LEVEL	7			SERVICEABLE_BALANCE	10	N
DEMILITARIZATION_CODE	1	A		SEX_CODE	1	A/N
DLA_STORAGE_FLAG	1			SHELF_LIFE_CODE	1	A/N
EEX_CODE	1	A/N		SPI_EFFECTIVE_DATE	DATE	
EQUIP_MGT_CODE	1	N		SPI_INDICATOR	1	
ERRCD	3	A/N		SPI_NUMBER	5	B
EXCESS_CAUSE_CODE	1	A/N		SRD_COLLECTION_FLAG	1	
FAST_TRANS_DENIAL_CODE	1	A/N		STANDARD_DEVIATION	1	N
FILE_STATUS_QUARTER_CODE	1	A/N		STOCK_FUND_CREDIT_FLAG	1	A/N
FILLER_3	9			STOCKAGE_PRIORITY_CODE	1	A/N
FILLER_4	4			SUPPLEMENTAL_ADJUNCT_RCD_FLAG	1	A/N
FILLER_5	1			SUPPLY_POINT_RCD_FLAG	1	A/N
FIXED_LEVEL_FLAG	1	A/N		SUSPECT_MATERIAL_FLAG	1	A/N
FOAM_IN_PLACE_FLAG	1	A		SYS_DESIG	2	A/N
FORECAST_ACQUISITION_COST	NUMBER(10,2)			TCTO_FLAG	1	A/N
FREEZE_CODE	1	A		TYPE_CARGO_CODE	1	A/N
FULLY_INTERCHANGEABLE_FLAG	1			TYPE_PROCUREMENT_CODE	1	A
FUNCTIONAL_CHECK_FLAG	1	A/N		TYPE_SRAN	1	A
HAZARDOUS_MATERIAL_CODE	1	A		UNIT_OF_ISSUE	2	A
HEALTH_HAZARD_FLAG	1	B		UNIT_PRICE	8	N
IEX_CODE	1	A/N		UNSUITABLE_ITEM_FLAG	1	
ISG_NUMBER	4			WARRANTY_CODE	1	
ISG_ORDER_CODE	2	A/N		XCE_DATE	4	

Appendix G – Data Element Definitions and Parameters

Acquisition Advice Code – indicates how and under what restrictions an item will be acquired. Also used to identify disposal, condemned, semi-active, and local purchase/local-manufacture items during supply decision processes

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
AAC_CD	ACQUISITION_ADVICE_CODE	1/alphabetic
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 58	V1, P4, Ch1, Table 1A54.2	A through Z

Automatic Data Processing Equipment (ADPE) Identification Codes – identifies DoD ADPE/ADP equipment and components in the supply system

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
ADP_EQP_ID_CD	ADPE_FLAG	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 159	V1, P4, Ch1, Table 1A57.1	0 through 9

Budget Code – identifies investment items to budget programs from which procurement of the particular item is funded, or to identify expense items to the various divisions of the Air Force Stock Fund

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
BUD_CD	BUDGET_CODE	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 67	V1, P4, Ch1, Table 1A42.1	A through Z, 1, 4, 6, 8, 9, @, *

Demilitarization Code – indicates if demilitarization is needed and how to carry it out

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
DEMIL_CD	DEMILITARIZATION_CODE	1/alphabetic
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 192	V1, P4, Ch1, Table 1A47.1	A through G, P, Q

Expendability, Recoverability, Reparability, Cost Designator (ERRCD) – used to categorize AF inventory into various management groupings

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
EXPND_RECVR_RPR_CD	ERRCD	3/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 69	V2, P2, Ch3, Table 3A5.1	XD1, XD2, XF3, XB3, ND, NF (for SBSS)
		C, T, P, N, S, U (for D043A)

ND/NF ***can* be followed by 1 through 5
**Used interchangeably between systems in the respective order above

Federal Supply Class (FSC) – identifies the commodity class of an item and appears in the first four positions of a stock number

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
FED_SUPL_CLASS_NR	FSC	4/numeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V4, Ch2, Para 4.2.1	V2, P2, Ch3, Para 3A1.2, Pg 21	4-digit numeric

Freeze Code – restricts processing of selected inputs, and identifies the activity responsible and the reason for freezing an item record

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
FRZ_CD	FREEZE_CODE	1/alpha
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
n/a	V2, P2, Ch27, Para 27.103.4.1 - 27.103.4.10	A, C, D, E, I, L, P, Q, R, S, empty

**assume codes are same across systems

Hazardous Materiel Identification Code (HMIC) – identifies items that require special handling, storage, use, transportation, and disposal because of hazardous materiel

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
HAZ_MTL_IND_CD	HAZARDOUS_MATERIAL_CODE	1/alpha
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 179	V2, P2, Ch3, Para 3A1.2, Pg 22	Y, D, P, N

Item Name/Nomenclature – identifies items in graphic and specific terms

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
ITM_NM	NOMENCLATURE	19/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Tables 20, 185, 1	V2, P2, Ch3, Para 3A1.2, Pg 33	Ranges from 19-32 positions

**regulations vary, 19 characters allotted with 32 character maximum

National Item Identification Number (NIIN) – serves to fix the identity of an individual item of supply and to distinguish it concisely and permanently from all other items

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
NAT_ITM_ID_NR	NIIN	9/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V1, Pg 1-1B-2	V2, P2, Ch3, Para 3A1.2, Pg 32	9 characters

**first two digits pre-determined based on guidance, see following table

National Codification Bureau code, first two digits of NIIN			
00	United States	23	Greece
01	United States	24	Iceland
11	NATO	25	Norway
12	Germany	26	Portugal
13	Belgium	27	Turkey
14	France	28	Luxembourg
15	Italy	29	Argentina
17	Netherlands	66	Australia
18	South Africa	98	New Zealand
21	Canada	99	United Kingdom
22	Denmark		

Price Validation Codes – indicates the validity of the recorded unit price

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
PRC_VAL_CD	PRICE_VALIDATION_CODE	1/alpha
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 177	V7, P4, Ch4, Table 4A1.1	A, D, E, N, P, V, X

Precious Metals Indicator Code (PMIC) – identifies items containing precious metals including gold, silver, and platinum

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
PREC_MET_IND_CD	PRECIOUS_METALS_FLAG	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 160	V6, Ch4, Table 4.1	A, C, G, P, S, U, V

Quantity Unit Pack Code (QUP) – indicates the number of Units of Issue in the unit package as established by the managing activity

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
QY_UNIT_PK_CD	QTY_UNIT_PACK_CODE	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 56	V1, P4, Ch1, Table 1A48.1	0 through 9, A through Z

**excluding “I” and “O”

Serialized Report Code (SRC) – indicates items designated as having characteristics that require they be identified, accounted for, secured, segregated, or handled in a special manner to ensure their safeguard or integrity

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
SER_RPT_CD	SERIALIZED_REPORT_CODE	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 61	V2, P2, Ch27, Att 27K-5	A through Z, 0 through 9, $, *

Shelf Life Code – indicates on the item record the number of months a new item may remain unused in storage before it must be reconditioned or condemned

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
SHLF_LIFE_CD	SHELF_LIFE_CODE	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 50	V2, P2, Ch3, Table 3A1.43	0 through 9, A through Z

**excluding "O"

Stock Fund Credit Flag/Code – identifies on the item record that credit will/will not be allowed for serviceable turn-ins

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
STK_FND_CR_CD	STOCK_FUND_CREDIT_FLAG	1/alphanumeric
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
n/a	V2, P2, Ch3, Para 3A1.2, Pg 50	A, D

Unit of Issue – codes/terms authorized for assignment to items of supply to identify unit of issue

System		
D043 Data Element	**SBSS Data Element**	**Size/Type**
UI_CD	UNIT_OF_ISSUE	2/alpha
Source		
DoD 4100.39-M	**AFMAN 23-110**	**Valid Fills**
V10, Table 53	V1, P4, Ch1, Table 1A6.1	see list

Unit of Issue - Valid Fills							
AM	BR	CO	GP	LT	PR	SH	TN
AT	BT	CD	GR	MC	PT	SK	TO
AY	BX	CY	HD	ME	PZ	SL	TS
BA	CA	CZ	HK	MM	QT	SO	TU
BE	CB	DR	IN	MR	RA	SP	VI
BF	CE	DZ	JR	MX	RL	SV	YD
BG	CF	EA	KG	OT	RM	SX	
BK	CK	FT	KT	OZ	RO	SY	
BL	CL	FV	LB	PD	SD	TD	
BD	CM	FY	LG	PG	SE	TE	
BO	CN	GL	LI	PM	SF	TF	

Appendix H – Completeness Results

	Valid	Invalid	% Valid	% Invalid	% of Total Invalid
FEDERAL_SUPPLY_CLASS	341,743	0	100.00%	0.00%	0.00%
NATIONAL_ITEM_IDENTIFICATION_NUMBER	341,743	0	100.00%	0.00%	0.00%
ACQUISITION_ADVICE_CODE	341,743	0	100.00%	0.00%	0.00%
ADPE_FLAG	274,466	67,277	80.31%	19.69%	9.64%
APPL DATA	2,530	339,213	0.74%	99.26%	48.59%
BUDGET_CODE	341,742	1	100.00%	0.00%	0.00%
DEMILITARIZATION_CODE	341,683	60	99.98%	0.02%	0.01%
ERRCD	341,743	0	100.00%	0.00%	0.00%
FREEZE_CODE	294,089	47,654	86.06%	13.94%	6.83%
HAZARDOUS_MATERIAL_CODE	316,058	25,685	92.48%	7.52%	3.68%
ITEM NAME/NOMENCLATURE	341,663	80	99.98%	0.02%	0.01%
PRICE_VALIDATION_CODE	341,729	14	100.00%	0.00%	0.00%
PRECIOUS_METALS_FLAG	341,724	19	99.99%	0.01%	0.00%
QTY_UNIT_PACK_CODE	341,743	0	100.00%	0.00%	0.00%
SERIALIZED_REPORT_CODE	341,743	0	100.00%	0.00%	0.00%
SHELF_LIFE_CODE	339,706	2,037	99.40%	0.60%	0.29%
STOCK_FUND_CREDIT_FLAG	125,656	216,087	36.77%	63.23%	30.95%
UNIT_OF_ISSUE	341,743	0	100.00%	0.00%	0.00%
Total Data Elements	**5,453,247**	**698,127**			
	88.65%	**11.35%**			

Total Records	**341,743**
Unique NIINs	**137,430**
Total Data Elements	**6,151,374**

Appendix I – Consistency Results

	Consistent	Inconsistent	% Consistent	% Inconsistent	% of Total Inconsistent
FEDERAL_SUPPLY_CLASS	810,859	666	99.92%	0.08%	0.02%
ACQUISITION_ADVICE_CODE	800,360	11,165	98.62%	1.38%	0.37%
ADPE_FLAG	787,476	24,049	97.04%	2.96%	0.80%
BUDGET_CODE	791,776	19,749	97.57%	2.43%	0.66%
DEMILITARIZATION_CODE	804,373	7,152	99.12%	0.88%	0.24%
ERRCD	0	811,525	0.00%	100.00%	27.16%
FREEZE_CODE	446,097	365,428	54.97%	45.03%	12.23%
HAZARDOUS_MATERIAL_CODE	369,829	441,696	45.57%	54.43%	14.78%
ITEM NAME/NOMENCLATURE	0	811,525	0.00%	100.00%	27.16%
PRICE_VALIDATION_CODE	793,977	17,548	97.84%	2.16%	0.59%
PRECIOUS_METALS_FLAG	808,833	2,692	99.67%	0.33%	0.09%
QTY_UNIT_PACK_CODE	811,002	523	99.94%	0.06%	0.02%
SERIALIZED_REPORT_CODE	809,973	1,552	99.81%	0.19%	0.05%
SHELF_LIFE_CODE	811,181	344	99.96%	0.04%	0.01%
STOCK_FUND_CREDIT_FLAG	339,210	472,315	41.80%	58.20%	15.81%
UNIT_OF_ISSUE	811,206	319	99.96%	0.04%	0.01%
Total Data Elements	9,996,152	2,988,248			
	76.99%	23.01%			

Total Records Compared	811,525
Unique NIINs	126,833
Total Data Elements	12,984,400

Excluding ERRCD and NOMENCLATURE					
	Consistent	Inconsistent	% Consistent	% Inconsistent	% of Total Inconsistent
FEDERAL_SUPPLY_CLASS	810,859	666	99.92%	0.08%	0.05%
ACQUISITION_ADVICE_CODE	800,360	11,165	98.62%	1.38%	0.82%
ADPE_FLAG	787,476	24,049	97.04%	2.96%	1.76%
BUDGET_CODE	791,776	19,749	97.57%	2.43%	1.45%
DEMILITARIZATION_CODE	804,373	7,152	99.12%	0.88%	0.52%
FREEZE_CODE	446,097	365,428	54.97%	45.03%	26.77%
HAZARDOUS_MATERIAL_CODE	369,829	441,696	45.57%	54.43%	32.35%
PRICE_VALIDATION_CODE	793,977	17,548	97.84%	2.16%	1.29%
PRECIOUS_METALS_FLAG	808,833	2,692	99.67%	0.33%	0.20%
QTY_UNIT_PACK_CODE	811,002	523	99.94%	0.06%	0.04%
SERIALIZED_REPORT_CODE	809,973	1,552	99.81%	0.19%	0.11%
SHELF_LIFE_CODE	811,181	344	99.96%	0.04%	0.03%
STOCK_FUND_CREDIT_FLAG	339,210	472,315	41.80%	58.20%	34.60%
UNIT_OF_ISSUE	811,206	319	99.96%	0.04%	0.02%
Total Data Elements	9,996,152	1,365,198			
	87.98%	12.02%			

Total Records Compared	811,525
Unique NIINs	126,833
Total Data Elements	11,361,350

Appendix J

Captain Craig A. Lane, Student, AFIT
craig.lane@us.af.mil
word count: 694

Data Quality – A Key to Successfully Implementing ECSS

In response to the Expeditionary Logistics for the 21st Century (eLog21) campaign initiatives published in 2003, the United States Air Force (USAF) pursued the acquisition of technology to help transform its logistics processes. With process mapping complete and a proposed roll-out schedule, forward progress towards full implementation of the Expeditionary Combat Support System (ECSS) continues. As a key enabler to achieving eLog21 initiatives, implementing ECSS will help transform current USAF logistics business processes. ECSS is the largest enterprise resource planning (ERP) system implementation in the world. When fully operational capability is reached, ECSS will have integrated several hundred legacy systems, and will serve in excess of 750,000 primary, secondary, and tertiary users. While the driving force behind an ERP system implementation is exploitation of the numerous benefits associated with transforming business processes, there are several key challenges to address which can mean the difference between success and failure. Data quality is one success factor consistently identified in literature as a critical part of any successful ERP system implementation. Quality data is a pivotal to optimizing system performance while maintaining an uninterrupted and acceptable level of support to the war fighter.

The literature on the subject spans both the military and the commercial sectors. Two key themes are consistent: the importance of data quality to a successful ERP implementation as well as the need to cleanse data prior to any ERP system implementation. However, a large gap exists regarding how and/or where to focus data cleansing efforts. I recently finished a study that focused on two legacy systems, one of which is slated to be a data source for ECSS, and found that current data residing in those systems was less than perfect. My study also identified a lack of any standard across the USAF with regard to data terminology or how quality data is defined. The results of the study identified data elements with invalid entries and highlighted 3 data elements which were the highest drivers of invalid data.

The existing processes that will eventually be absorbed by ECSS have been mapped and blueprinted to ensure they will be accurately carried forward into the new system. At this point, it appears there is no plan in place to do the same for the data being migrated into ECSS. There should be. I researched data quality, focusing on the completeness and consistency of the data, in selected USAF legacy systems. Specifically, my study identified invalid entries in the source data and also compares item record data between source (D043A) and downstream client (SBSS). My study revealed several important lessons which should be applied to the data being used to populate ECSS. First, the existing data was proven to be less than perfect. Second, my research identified the need to map individual data elements vice entire records. Third, I was able to identify data elements which appear to have the highest percentages of invalid entries.

This provides a foundation to sample the data in other legacy systems. Additionally, it identifies areas to focus cleansing efforts in a resource-constrained operating environment.

If the future state of the Air Force logistics enterprise hinges on current transformation efforts, then the successful implementation of ECSS is a critical piece of the success puzzle. Furthermore, quality data is necessary to exploit the benefits of ECSS to the fullest extent as well as optimize its performance. The USAF is investing a significant amount of tax-payer dollars, in excess of $1 billion, into the development and implementation of ECSS. This amount dwarfs the cost of most aircraft in our inventory. As a prior-enlisted POL troop, I'm certain that leadership would not condone refueling any aircraft with less than perfect fuel. This same logic should be applied to ECSS regarding data. Data quality is a real concern at this point, prior to the implementation of ECSS. This is the time to apply the proper resources to the appropriate data to address cleansing efforts and mitigate inaccuracies, before data is moved into the new system. In the information technology arena, it is widely accepted that garbage in equals garbage out. As the old adage goes, "an ounce of prevention is worth a pound of cure".

Craig Lane is a student at the Air Force Institute of Technology.

Bibliography

Air Force Materiel Command (AFMC). *Cataloging and Standardization.* AFMCMAN 23-3. Wright-Patterson AFB: HQ AFMC, 19 January 2007.

Becker, Dave, 554th Electronic Systems Wing. "Enterprise Data Quality Management for Operational Support." Briefing. Air Force Materiel Command, Wright-Patterson AFB, OH. 8 January 2009.

Bose, I., Pal, R., & Ye, A. (2008). ERP and SCM Systems Integration: The Case of a Valve Manufacturer in China. *Information & Management*, *45*(4), 233-241.

BusinessDictionary.com. *Definition of ERP*. http://www.businessdictionary.com/definition/enterprise-resource-planning-ERP.html. 6 February 2009.

Caruso, D. (2007). Six Ways to Ensure an ERP Implementation Delivers Value. *Manufacturing Business Technology*, *25*(8), 27.

Coker, David W. (2006). Lessons Learned from the Army's Largest ERP Implementation. *Defense AT&L*. Nov-Dec, 8-11. 20 May 2008 http://www.dau.mil/pubs/dam/11_12_2006/11_12_2006_cok_nd06.pdf

Defense Logistics Agency (DLA). *Transformation Roadmap*. Fort Belvoir, Virginia, October 2005.

Department of the Air Force (DAF). *Expeditionary Logistics for the 21st Century Campaign Plan*. Washington: HQ USAF, 2003.

Department of the Air Force (DAF). *USAF Supply Manual.* AFMAN 23-110. Washington: HQ USAF, 1 January 2009. http://www.e-publishing.af.mil/shared/media/epubs/afman23-110.pdf

Dredden, Glenn and Jeffrey C. Bergdolt. "Enterprise Resource Planning," *Air Force Journal of Logistics*, *31*(2), 48-52 (Summer 2007).

Falvey, David J. Leap Forward. *Fedtech Magazine.* 20 May 2008 http://fedtechmagazine.com/article.asp?item_id=355

Fawcett, S.E.; Ellram, L.M. and Ogden, J. A. (2007). Supply Chain Management: From Vision to Implementation. Prentice Hall, Upper Saddle River, NJ.

Fleisch, E., & Tellkamp, C. (2005). Inventory inaccuracy and supply chain performance: A simulation study of a retail supply chain. *International Journal of Production Economics*, *95*(3), 373.

Fri, Daniel A. "Enterprise Architecture," *Air Force Journal of Logistics, 31*(2), 40-46 (Summer 2007).

Grantner, Emily. "ISO 8000 – A Standard for Data Quality," *Logistics Spectrum, 41*(2), 4-6 (October-December 2007).

Harris, J. (2003). Anticipating the Risks of New Systems. *Risk Management*, *50*(7), 28.

Judson, David and Kinney, Bob. Vice President and Associate, Qbase™, Dayton OH. Personal Correspondence/Electronic Message. 12 February 2009.

Kumar, V., Maheshwari, B., & Kumar, U. (2003). An Investigation of Critical Management Issues in ERP Implementation: Emperical Evidence from Canadian Organizations. *Technovation, 23*(10), 793-807.

Lawrence, Barry F., Jennings, Daniel F., & Reynolds, Brian E. *ERP in Distribution*. Mason: South-Western, 2005.

Mabert, V. A., Soni, A., & Venkataramanan, M. A. (2003). Enterprise Resource Planning: Managing the Implementation Process. *European Journal of Operational Research, 146*(2), 302-314.

Ngai, E. W. T., Law, C. C. H., & Wat, F. K. T. (2008). Examining the Critical Success Factors in the Adoption of Enterprise Resource Planning. *Computers in Industry, 59*(6), 548-564.

Pugh, Don, HQ AFMC/A4N. "Expeditionary Combat Support System (ECSS) Transforming AF Logistics." Address to ECSS Roadshow audience. Wright-Patterson AFB OH. 22 May 2008.

Scott Schumacher (2007, November). Risk Management and the Fog of Data. *The Secured Lender, 63*(6), 80,82. Retrieved May 25, 2008, from ABI/INFORM Research database. (Document ID: 1392423881).

Sun, A. Y. T., Yazdani, A., & Overend, J. D. (2005). Achievement Assessment for Enterprise Resource Planning (ERP) System Implementations Based on Critical Success Factors (CSFs). *International Journal of Production Economics, 98*(2), 189-203.

Tersine, Richard J. *Principles of Inventory and Materials Management, 4th ed.* New Jersey: Prentice Hall, 1994.

Tew, Scott. Global Logistics Support Center Master Briefing, 12 December 2006.

Titmuss, K. (2001). “Inventory Accuracy, The First Step to Successful Supply Chain Management”, paper presented at the South African Production and Inventory Control Society Conference, Durban, South Africa. 25 July 2008 http://www.kos.co.za/InvAcc/InvAcc.htm

Titmuss, K. (2007). “The ERP Delinquent Dozen: Why ERP Systems rarely provide most Manufacturing Companies with a Competitive Advantage in their Supply Chain,” *Proceedings of the SAPICS 29th Annual Conference and Exhibition.* 1-9. SAPICS 2007. 15 January 2009 http://www.kentoutserv.com/ERPDelinquentDozen.pdf

Trunik, P.A. (1999), “ERP: promise or pipe dream?”. *Transportation & Distribution,* Vol. 40 No. 1, pp. 23-26.

Umble, E. J., Haft, R. R., & Umble, M. M. (2003). Enterprise Resource Planning: Implementation Procedures and Critical Success Factors. *European Journal of Operational Research, 146*(2), 241-257.

U.S. Department of Defense (DoD). *2007 Enterprise Transition Plan (ETP).* 28 September 2007. 5 February 2009 http://www.pentagon.gov/dbt/products/ETP_Excerpts/ExtractSept07.pdf

U.S. Department of Defense (DoD). *Federal Logistics Information System(FLIS) Procedures General/Administrative Information*. DoD 4100.39-M, Volume 1. Washington: HQ USAF, February 2009.

U.S. Department of Defense (DoD). *Federal Logistics Information System(FLIS) Procedures Manual Item Identification*. DoD 4100.39-M, Volume 4. Washington: HQ USAF, February 2009.

U.S. Department of Defense (DoD). *Federal Logistics Information System(FLIS) Procedures Manual Multiple Application References/Instructions/Tables and Grids*. DoD 4100.39-M, Volume 10. Washington: HQ USAF, February 2009.

Yu, C. (2005). Causes Influencing the Effectiveness of the Post-Implementation ERP System. *Industrial Management and Data Systems, 105*(1/2), 115.

Yun Kang, & Gershwin, S. B. (2005). Information Inaccuracy in Inventory Systems: Stock Loss and Stockout. *IIE Transactions, 37*(9), 843-859.

Vita

Captain Craig A. Lane graduated from Holley High School in Holley, New York. He enlisted in the USAF and entered active duty on 9 May 1994, as a Fuels Specialist. His first assignment was the 325th Fighter Wing at Tyndall AFB, Florida, where he served as a refueling unit operator and preventative maintenance technician. In December 1997, he was reassigned to the 82nd Training Wing at Sheppard AFB, Texas, as a Technical Training Instructor. While stationed at Sheppard, he enrolled in the Occupational Education undergraduate program through Wayland Baptist University. In 2000, he graduated with a Bachelor of Science degree and immediately enrolled in their graduate program. In 2001, he graduated with a Master of Arts in Management and was subsequently accepted to attend Officer Training School. He received his commission on 20 December 2001, and was assigned to Minot AFB, North Dakota, where he served primarily as the Installation Deployment Officer. While stationed at Minot, he deployed to Guam in support of Operations ENDURING and IRAQI FREEDOM (OEF/OIF). In December 2004, he was reassigned to Ellsworth AFB, South Dakota, where he served as the Materiel Management Flight Commander and interim Squadron Operations Officer. In October 2005, he deployed to Kandahar, Afghanistan. While deployed, he served as part of a 5-man Embedded Tactical Training Team, mentoring the Afghan National Army (ANA) on logistics. In April 2007, he deployed to US Central Command and served on the J4 staff, Logistics Automation Division. In August 2007, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology. Upon graduation with a Masters Degree in Logistics Management, he will be assigned to the Air Force Logistics Management Agency, Gunter Annex, Maxwell AFB, AL.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
18-03-2009	Master's Thesis	Sep 2007- Mar 2009

4. TITLE AND SUBTITLE

DATA QUALITY – A KEY TO SUCCESSFULLY IMPLEMENTING ECSS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Lane, Craig A., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENS)
2950 Hobson Street, Building 642
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GLM/ENS/09-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Logistics Transformation Office
Attn: Mr. Steven Cain
4375 Chidlaw Rd Bldg 262 DSN: 785-4539
WPAFB OH 45433-7765 e-mail: steven.cain@wpafb.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In response to the Expeditionary Logistics for the 21st Century (eLog21) campaign initiatives published in 2003, the United States Air Force (USAF) pursued the acquisition of the technology needed to transform its logistics processes. Today, with process mapping complete and a firm base release schedule, forward progress towards full implementation of the Expeditionary Combat Support System (ECSS) continues. As a key enabler to achieving eLog21 initiatives, implementing ECSS will completely transform current USAF logistics business processes. Integrating more than 450 legacy systems, and with a projected end-state in excess of 750,000 primary, secondary, and tertiary users, ECSS is the largest enterprise resource planning (ERP) implementation in the world. While the driving force behind ERP implementation is exploitation of the numerous benefits associated with transformation, there are several key challenges to address which can mean the difference between success and failure. Data quality is a critical factor in the successful implementation of any ERP system. It is a key factor in optimizing system performance and maintaining an uninterrupted and acceptable level of support to the war fighter. The following research provides a representative snapshot and evaluates current item record data quality in selected USAF legacy systems; specifically a comparison of data in the source system (D043A) versus the same data in the client system (SBSS). The analysis lays the foundation for developing an action plan for efficient and effective resource allocation to cleanse the legacy system data prior to migrating into ECSS, thereby creating an operating environment more prone to a successful implementation.

15. SUBJECT TERMS

Data, Data Quality, ERP Implementation, Expeditionary Combat Support System (ECSS), Transformation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Jeffrey A. Ogden, PhD (ENS)
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	95	937-255-3636

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18